2012

Transcriptomic identification and characterization of levamisole resistance associated genes in the swine nodular worm Oesophagostomum dentatum

Nathan Romine
Iowa State University

Transcriptomic identification and characterization of levamisole resistance associated genes in the swine nodular worm *Oesophagostomum dentatum*

by

Nathan Michael Romine

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Genetics (Computational Molecular Biology)

Program of Study Committee:
Jeffrey K. Beetham, Co-major Professor
Hui-Hsien Chou, Co-major Professor
Adam J. Bogdanove
Douglas E. Jones
F. Chris Minion

Iowa State University

Ames, Iowa

2012

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Treatment of parasitic nematodes infections is generally limited to one of three major drug classes; resistance to these is an increasing problem. Because development of new drugs and drug classes is expensive and slow it is important to understand how resistance to current drugs occurs. Nicotinic acetylcholine receptors provide drug targets for both the nicotinic agonist and amino-acetonitrile derivative anthelminthic classes. Much of the research on resistance to nicotinic agonists has been performed in the free-living nematode *Caenorhabditis elegans*. This research has by necessity looked at limited gene sets which present an incomplete picture of what is believed to be a polygenic trait. Attempts to reproduce this research in parasitic species have shown that research in *C. elegans* does not always translate to parasites.

We used second generation sequencing to obtain a broader view of resistance in a parasitic nematode, *Oesophagostomum dentatum*, than is easily accomplished with traditional molecular methods. Because *O. dentatum* lacks a sequenced genome it was first necessary to identify mRNA sequences for genes shown in other genera/species to associate with resistance. We developed a method of assembly that produces longer sequences than traditional assembly methods and used that to identify mRNAs for 34 genes associated with resistance to levamisole and other major anthelminthics.

With this sequence information we assessed the expression levels and sequence changes in the levamisole resistance associated genes between levamisole-sensitive and -resistant nematode isolates. We identified 9 mRNAs exhibiting at least a 2-fold decrease in expression between the two isolates and 72 non-synonymous SNPs. We have used this information to propose that levamisole resistance in this parasitic model associates with decreased abundance of functional receptors containing UNC-38 and/or UNC-63, as well as decreased signal transduction moderated by LEV-10, NRA-1, RIC-3, LEV-11, UNC-22, and UNC-68_{ry}.

CHAPTER 1. GENERAL INTRODUCTION

Treatment of parasitic nematodes is generally limited to one of three classes of drugs; the benzamidizoles, macrolytic lactones, and nicotinic agonists. Resistance to the three major anthelminthic classes is becoming a major problem, increasingly complicating treatment of human infections and endangering the viability of raising livestock in some countries. Development of new anthelminthics has been slow; only two new anthelminthic classes have been introduced in the last 30 years and their commercial availability is limited. For this reason it is necessary to understand the mechanisms of resistance to current anthelminthics.

Nicotinic acetylcholine receptors have much greater diversity in nematodes than in mammals and provide targets for both the nicotinic agonist and amino-acetonitrile derivative anthelminthic classes. Much of the research on resistance to nicotinic agonists has been performed in the free-living nematode *Caenorhabditis elegans*. This research has by necessity looked at limited gene sets which present an incomplete picture of what is believed to be a polygenic trait. Additionally, several attempts to reproduce this research in parasitic species have shown that research in *C. elegans* does not always translate to parasites. The research detailed here presents a broader view of resistance in a parasitic nematode, *Oesophagostomum dentatum,* than is easily accomplished with traditional molecular methods.

The goal of this work is to improve treatment of parasitic nematode infections. Improved treatment techniques will benefit both health professionals and livestock producers.

Dissertation Organization

A brief introduction to this dissertation describes the problem of anthelminthic resistance and the limits of traditional molecular research. Chapter 2 reviews the literature relevant to the research system discussed in the two research chapters; this chapter is formatted for BMC Genetics. The first of the two research chapters (chapter 3) details a method used to identify target genes from mRNA-Seq data and is formatted for submission to the Journal of Parasitology. The second of the research

chapters (chapter 4) uses the method described in chapter 3 to identify genes that are part of the nicotinic acetylcholine receptor pathway and evaluate them for sequence or expression changes between levamisole-sensitive or -resistant isolates to identify molecular mechanisms of levamisole resistance. Chapter 4 is formatted for submission to BMC Genetics. The final chapter is a general discussion of the results of the research performed for this dissertation. This chapter also explores some of the research being continued from these studies or made possible by this work. The discussion chapter is also formatted for BMC Genetics.

CHAPTER 2. LITERATURE REVIEW

Nematodes

The phylum nematoda is one of the largest in the animal kingdom second only to arthropoda; estimates put the number of nematodes species around 1,000,000. Nematoda is also one of the most populous phyla accounting for nearly 80% of the multicellular animals on Earth [1]. Nematodes have adapted to marine and terrestrial environments as well as both polar and tropical climates. Nematodes as a phylum have significant biological importance. Some species play a significant role in the decomposition of matter in soil. *Caenorhabditis elegans*, one of the most widely used experimental systems and the first multicellular organism to have its genome sequenced, is a member of the phylum. Furthermore, a large proportion of identified species are parasites of plants (10%) or animals (15%).

Nematodes are roundworms with a tubular digestive system that is open on both ends. They possess a collagenous cuticle covering their epidermis with a layer of muscle beneath the epidermis. Nematodes also have a muscular pharynx near the anterior end and a sphincter muscle controlling the rectal opening. The nervous system of the nematode *C. elegans* is one of the most understood systems in biology and is a usable model for other nematodes. A bundle of nervous tissue near the pharynx of *C. elegans* acts as a brain. From this central nervous system radiates a dorsal and a ventral nerve extending to either end of the worm. Additional neurons radiate laterally from the central nerves.

Nicotinic Acetylcholine Receptors

Acetylcholine (ACh) is an excitatory neurotransmitter and is the primary neurotransmitter at nematode neuromuscular junctions. ACh receptors (AChR) are identified pharmacologically by their major agonist, forming two types of receptors: the muscarinic and the nicotinic ACh receptors. Each nicotinic AChR (nAChR) is a heteropentameric ligand-gated ion channel composed of multiple subunits. Gene prediction of the *C. elegans* genome has shown that there might be as many as 50

genes that code for nAChR subunits [2]. Jones and Satelle [3] identified 27 expressed nAChR subunit genes in *C. elegans* which may be categorized based on homology into 5 families: *acr-8*, *acr-16*, *deg-3*, *unc-29*, and *unc-38*. This gives *C. elegans* the most diverse set of recognized nAChR subunit genes. Jones and Satelle [3] further characterized the *acr-8* and *deg-3* families as nematode-specific families. These subunits are additionally classified as alpha subunits and non-alpha subunits depending on their biological function.

Studies of vertebrates have shown that nAChRs are arranged in a circle at the cell membrane, forming a pore (Figure 2.1). These nAChRs are formed of two or more alpha subunits and three or fewer non-alpha subunits. Ligand binding occurs at the junction of the alpha subunit with the adjacent subunit [4, 5]. Each of the nAChR subunit families has varying numbers of alpha and non-alpha subunits. The acr-8 family has three alpha subunits while the acr-16 family has seven alpha and two non-alpha subunits. The deg-3 family has seven alpha and one non-alpha subunit while the unc-29 family has 4 non-alpha subunits and the unc-38 family has 3 alpha subunits.

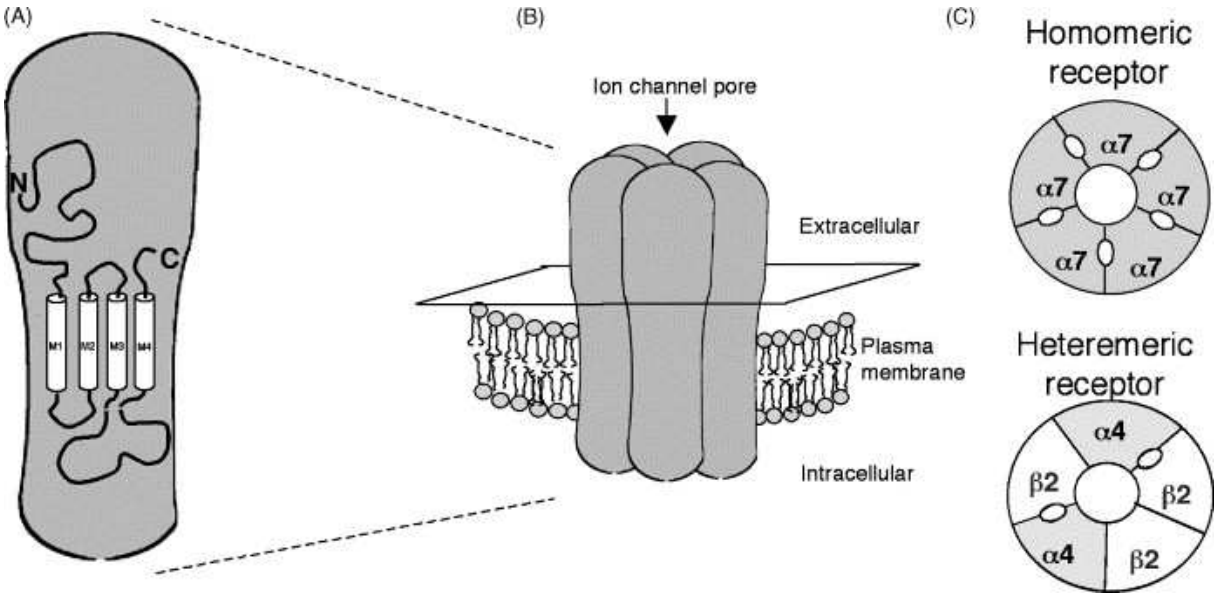

Figure 2.1 - Structure of the nAChR. nAChRs are pentameric receptors that create a pore through the cell membrane. Various combinations of receptor subunits produce receptors with different properties. Reprinted with permission from Gotti and Clementi [6]

Soil-transmitted Helminthiases

Soil-transmitted helminthiases (STH) are nematode infections that represent the most prevalent of the neglected tropical diseases. Recent estimates suggest that more than one billion people are infected with one or more STH infections [7, 8]. The route of infection is dependent on the species of parasite with some infecting a host through the ingestion of eggs or larvae from contaminated soil while larvae of other species penetrate the skin after contact with soil. Most STH infections occur in sub-Saharan Africa, Latin America, China, and East Asia. The morbidity of STH is profound and estimated to be 4.7 million disability adjusted life years [9], although King [10] warns that the WHO is underestimating the incidence by not including the full clinical spectrum of STH symptoms. Symptoms of STH include malaise and weakness; intestinal infections may result in malnutrition and anemia. Children are most affected by STH infection, with resultant growth retardation and educational delays. Severe infections can be fatal, with an estimated 155,000 deaths per year attributed to STH infection [11]. In addition to the impact on human heath, helminths are of agricultural concern. Various helminth species infect economically important plants and animals reducing agricultural production and causing as much as US $110 billion in economic loss [12].

Oesophagostomum species are soil-transmitted intestinal helminths with a life cycle characteristic for the STH (Figure 2.2). Various *Oesophagostomum* species infect cattle, goats, swine, humans, and non-human primates. Oesophagostomiasis, the disease resulting from *Oesophagostomum* infection, is transmitted by consumption of soil contaminated with *Oesophagostomum* larvae. Larvae then burrow in to the wall of the large intestine where they form a nodule and mature. Larvae move back to the lumen of the colon to mature to adults and copulate. Eggs are shed in the feces where they develop through several stages of larvae into the infectious stage.

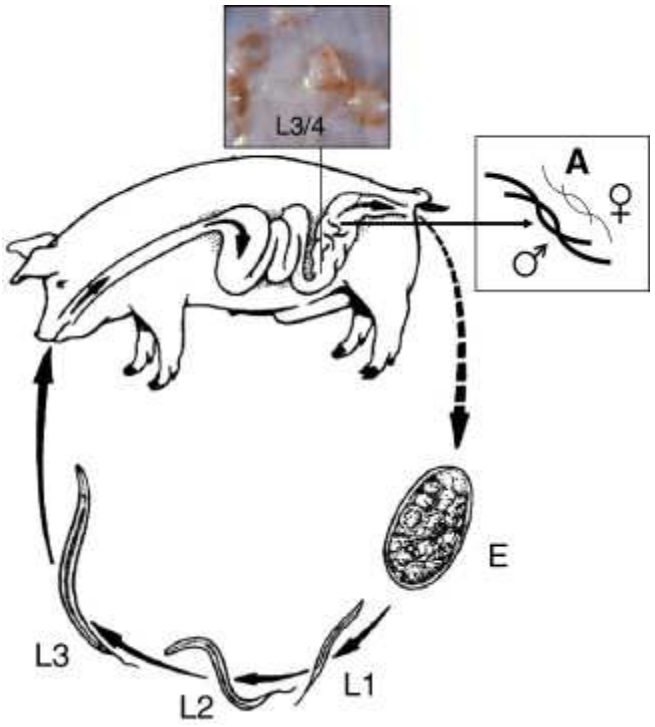

Figure 2.2 - Life cycle of *O. dentatum*. L3 larvae are ingested by a pig. These larvae imbed in the host intestine to develop into adults. Adults move back to the lumen of the intestine to reproduce. Eggs are shed in the feces and develop in the environment. Reprinted with permission from Gasser [13]

Anthelminthics

Treatments for helminth infections have been in development since the dawn of medicine. The treatment for guinea worm, winding the worm around a stick after it bursts through the skin, dates back to 1550 BC and is proposed to be the origin of the rod of Asclepius medical symbol [14]. In modern medicine, however, the most effective broad spectrum treatment is drug therapy. Five classes of anthelminthic drugs, categorized by mode of action (MOA), exist: benzimidazoles, macrolytic lactones, nicotinic agonists, cyclooctadepsipeptides, and amino-acetonitrile derivatives. Benzimidizoles were the first of the modern anthelminthic drugs when thiabendazole was introduced in 1961. Nicotinic agonists followed with the introduction of levamisole in 1970. Macrolytic lactones were the last of the major anthelminthic classes commercialized with the advent of ivermectin in 1981. Cyclooctadepsipeptides and amino-acetonitrile derivatives are the most recently introduced

classes of drugs, being first reported in 1992 [15] and 2008 [16]; however, they are not broadly available. As safety and efficacy testing for animal populations is less stringent, drugs are typically developed for animal use prior to being approved for use in humans. For this reason anthelminthics used for treatment of human infection are limited, although the WHO Model List of essential medicines lists several including: albendazole, levamisole, mebendazole, niclosamide, praziquantel, and pyrantel [17].

Drug Resistance

Development of resistance to the sparse catalog of anthelminthics is an area of concern. Drug resistance in general arises through one of four mechanisms: reduced drug accumulation, drug inactivation, alteration of target, or alteration of metabolic pathways. Drug accumulation may be lessened by diminished import, through alteration of pores through which drugs enter the cell, or by increased export of the drug via an efflux pump. Inactivation may occur through either metabolism of the drug, such as the deactivation of penicillin by β-lactamases in bacteria, or by no longer converting a prodrug to the active form, which is a common mechanism seen in cancer cells. Microbes employ strategies for alteration of the drug target. Alteration of the drug target may occur through overproduction of the target (overproduction of para-amino benzoic acid is a mechanism of resistance to sulfonamides) or outright elimination of the target. Alteration of the target may also occur through modification of the target's affinity for the drug, such as alteration of the penicillin-binding protein, another method of penicillin resistance. The final mechanism of drug resistance is through modification of the metabolic pathway on which the drug acts. This is accomplished by bypassing the site of drug action. For example, some bacteria have developed methods of taking up rather than synthesizing folic acid as a mechanism of sulfonamide resistance.

Given that drugs within a class have similar MOAs resistance to a drug is usually an early indicator of resistance to the entire class. Anthelminthic resistance has been observed in agricultural parasites for the three major anthelminthic classes benzamidazoles, macrolytic lactones, and nicotinic

agonists [18]. Figure 2.3 shows the timeline of drug introduction to first report of resistance. Field resistance has not been reported for the cyclo-octadepsipeptide and amino-acetonitrile derivative classes of drugs, although each has major limitations on commercial availability and approval for application. Data show that treatment regimens of human parasitic worm infections are increasingly failing which suggests that resistance is also mounting in human parasites [19, 20]. Because of the slow pace of drug development and the rapidity with which anthelminthic resistance develops it is important to understand how resistance occurs.

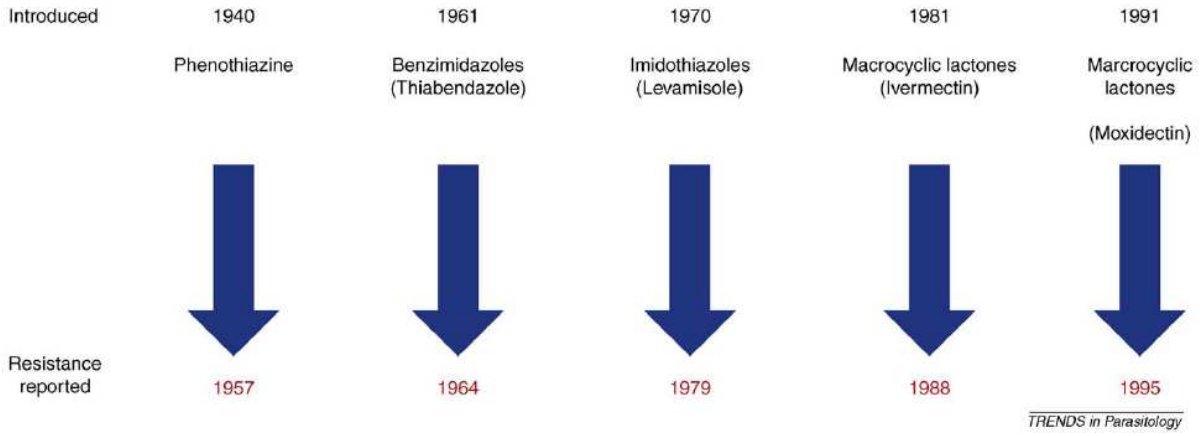

Figure 2.3 – Timeline of introduction to first report of resistance of anthelminthics. Reprinted with permission from James [21]

Nicotinic agonists act on the nAChR or associated proteins; given the wide diversity of nematode nAChR subunits and receptors these drugs still maintain specificity for the parasite. Levamisole is a broad spectrum nicotinic acetylcholine receptor (nAChR) agonist that has been critical for treatment of helminth infection of humans and animals. Levamisole has been shown to be more potent than the natural ligand, ACh [22-25] Levamisole activation of the nematode nAChR results in somatic muscle activation and spastic paralysis leading to the worm being expelled from the host gut by normal peristalsis. Field resistance of levamisole, first reported in 1979 [26], has become an increasing problem.

Lewis et al. [27] first used mutagenesis studies in *C. elegans* to identify a set of genes that could confer resistance to levamisole. Fleming et al. [28] demonstrated that several of these genes encoded nAChR subunits. Three of these were shown to associate with high resistance to levamisole, *unc-29* (non-alpha subunit), *unc-38* (alpha subunit), and *unc-63* (alpha subunit), while *lev-1* (non-alpha subunit) and *lev-8* (acr-13; alpha subunit) associate with low resistance to levamisole [29-31]. In addition to these receptor subunits a wide variety of proteins have been shown to associate with levamisole resistance at all levels of signal transduction (Figure 2.4).

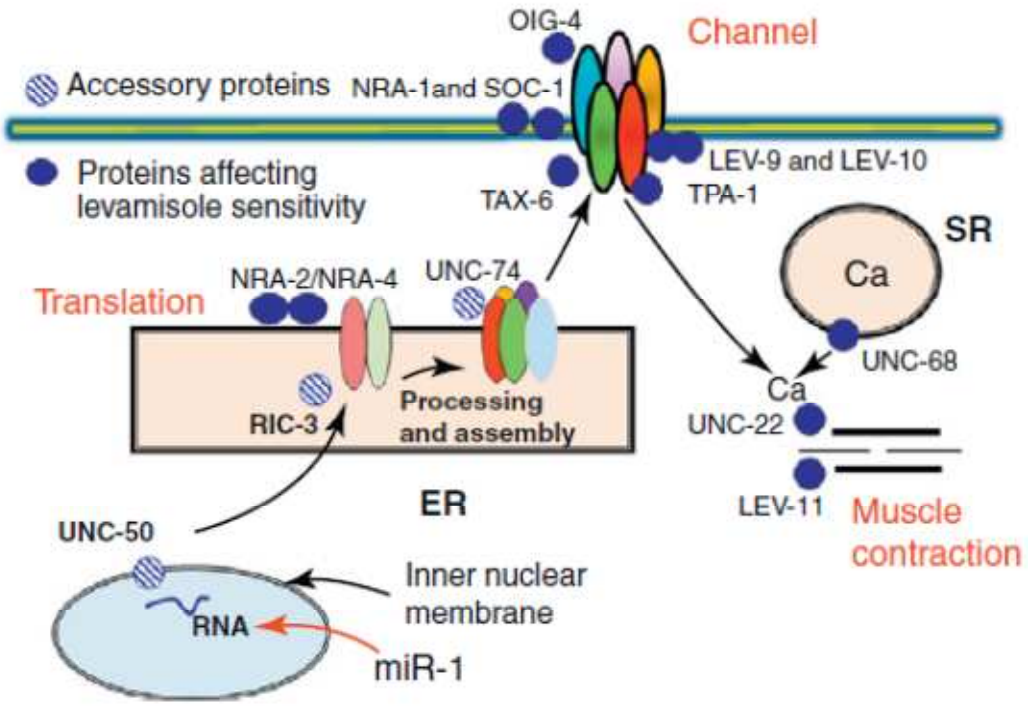

Figure 2.4 - Proteins shown to associate with levamisole resistance in *C. elegans*. Proteins involved in all aspects of nAChR formation and signal transduction have been show to associate with some level of levamisole resistance. Reprinted with permission from Martin et al. [41]

UNC-74 is a thioredoxin protein that appears to function solely in expression of levamisole receptor subunits [32]. NRA-2 and NRA-4 [33] and UNC-50 [34] affect composition of the receptor by directing subunit export from the Golgi; these subunits are either targeted to the cell surface or lysosomes for recycling. Several other proteins, including RIC-3 [35], NRA-1 and SOC-1 [36], and

LEV-10 [37], function in levamisole receptor formation at the cell surface. LEV-9 is a secreted protein that complexes with LEV-10 to localize receptors to the synapse [38]. OIG-4 is a secreted protein containing a single immunoglobulin domain and interacts with the LEV-9/LEV-10 complex [39]. TAX-6 encodes a phosphatase that negatively regulates the levamisole receptor [36]. TPA-1 encodes a protein kinase C homolog that regulates receptor function [3]. UNC-68 is a receptor on the sarcoplasmic reticulum which releases calcium to amplify the signal [40]. UNC-22 and LEV-11 encode twitchin and tropomysin, respectively, which are involved in muscle contraction [28].

Several genes have also been identified that affect benzimidazole or ivermectin resistance. Benzimidazoles act by binding to and disrupting microtubules. Logically, mutations in tubulin would be a mechanism of drug resistance, and indeed three mutations in beta tubulin can cause resistance: Phe200Tyr [42], Phe167Tyr [43], and Glu198Ala [44]. Macrolytic lactones activate the chloride channel gamma-aminobutyric acid receptors leading to neuromuscular paralysis. A Leu256Phe mutation in some of these chloride channels has been shown to cause resistance in *H. contortus* [45] although it is not sufficient for macrolytic lactone resistance in C. elegans [46].

Although *C. elegans* has proven to be a useful neurologic and genetic model, Kohler [47] warns of the pitfalls of studying resistance in non-parasitic model organisms. Because non-parasites, such as *C. elegans,* do not experience the stresses associated with surviving in a host, conclusions drawn from these studies may not be applicable to parasite populations. [48] provide additional evidence of the necessity of translating research in *C. elegans* to parasites: they were unable to identify the presence of *lev-8* in *H. contortus*, a pathogen of ruminants. As a parasite, *O. dentatum* is a more applicable research target than *C. elegans*.

Electrophysiological studies of levamisole resistance in *O. dentatum* have supported the role of the nAChR pathway in the development of levamisole resistance, as has been suggested in *C. elegans*. Robertson et al. [49] demonstrated that levamisole-resistant (LEVR) isolates had a decreased responsiveness to increasing levamisole concentrations as compared to a levamisole-

sensitive isolate (SENS). This decrease in receptor responsiveness could result from changes to the levamisole receptor subunit population, either subunit sequence or stoichiometry, or changes in subunit post-translational modifications. While it is generally accepted that anthelminthic resistance is polygenic [49-51], Kohler [47] suggests that receptor loss or reduced target affinity are the primary mechanisms of anthelminthic resistance.

Second Generation Sequencing

Although *O. dentatum* is a preferable study target given its role as a parasite, the challenge in studying resistance in non-model organisms is the lack of specific molecular tools. RNA-Seq provides a global measurement of transcript abundance [52], and in non-model species can help overcome the lack of specific molecular tools. RNA-Seq is performed by making cDNA from a fragmented mRNA population and using second generation sequencing (SGS) technology, such as Illumina sequencing, to obtain large volumes of data. The simplest method of interpreting RNA-Seq data involves mapping the transcripts to a genome. However, when no genome sequence is available it is necessary to de novo assemble a transcriptome. Although there are many difficulties associated with de novo transcriptome assembly including differential expression of transcripts, uneven coverage across a transcript, expression of transcripts from overlapping loci, and alternative splicing of transcripts [53] several programs for de novo transcript assembly are freely available [53-55]. Once a transcriptome is assembled there are many options available for assessment including analysis of normalized expression [56] and sequence changes [57, 58] between samples. Specific mRNAs of interest may also be identified from the transcriptome and for molecular assessments.

Earlier attempts at transcriptome assembly in *O. dentatum* yielded primarily short contigs [59]. The method used for that transcriptome assembly was not optimized for assembly of full length contigs. Mizrachi et al. [60] used paired end sequences for a transcriptome assembly and showed that mapping reads to contigs from a de novo assembly, collecting those reads, and reassembly could

correct for errors in the initial assembly. Work on SGS genome assemblies produced varied levels of improvement of assembly length by using paired reads to close gaps [61].

Although Robertson et al. [49] provided evidence that levamisole resistance occurs through modification of the nAChR of *O. dentatum*, identifying the basis for anthelminthic resistance is complicated by the polygenic nature of resistance. Studies that have explored resistance by traditional genetics methods have by necessity looked at a limited set of genes. SGS and computational analysis provide the first tools to look at resistance globally. Earlier attempts at transcriptomic assembly in *O. dentatum* produced an incomplete picture, but newer work has suggested methods of improving assembly. An improved transcriptomic assembly will facilitate RNA-Seq analysis of levamisole-sensitive and –resistant *O. dentatum* isolates. This, in turn, will aid in characterizing the molecular mechanisms for the differences observed between these two populations. Transcriptomic data from RNA-Seq will provide the sequence information necessary to define the nAChR pathway in *O. dentatum* and aid identification of levamisole targets. RNA-Seq analysis will also be used to determine an association between levamisole resistance and isolate dependent changes in protein sequence or stoichiometry. More information about the nAChR subunits will also lead to increased understanding of how different drugs act on different receptor subtypes.

References

1. Platt HM: **Foreward**. In: *The phylogenetic systematics of free-living nematodes.* Edited by Lorenzen S. London: The Ray Society; 1994.
2. *C. elegans* Sequencing Consortium: **Genome sequence of the nematode C. elegans: a platform for investigating biology**. *Science* 1998, **282**(5396):2012-2018.
3. Jones AK, Sattelle DB: **Functional genomics of the nicotinic acetylcholine receptor gene family of the nematode, Caenorhabditis elegans**. *BioEssays* 2004, **26**(1):39-49.
4. Corringer PJ, Bertrand S, Bohler S, Edelstein SJ, Changeux JP, Bertrand D: **Critical elements determining diversity in agonist binding and desensitization of neuronal nicotinic acetylcholine receptors**. *J Neurosci* 1998, **18**(2):648-657.
5. Brejc K, van Dijk WJ, Klaassen RV, Schuurmans M, van Der Oost J, Smit AB, Sixma TK: **Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors**. *Nature* 2001, **411**(6835):269-276.

6. Gotti C, Clementi F: **Neuronal nicotinic receptors: from structure to pathology**. *Prog Neurobiol* 2004, **74**(6):363-396.
7. Brooker S: **Estimating the global distribution and disease burden of intestinal nematode infections: adding up the numbers--a review**. *Int J Parasitol* 2010, **40**(10):1137-1144.
8. Bethony J, Brooker S, Albonico M, Geiger SM, Loukas A, Diemert D, Hotez PJ: **Soil-transmitted helminth infections: ascariasis, trichuriasis, and hookworm**. *Lancet* 2006, **367**(9521):1521-1532.
9. van der Werf MJ, de Vlas SJ, Brooker S, Looman CW, Nagelkerke NJ, Habbema JD, Engels D: **Quantification of clinical morbidity associated with schistosome infection in sub-Saharan Africa**. *Acta Trop* 2003, **86**(2-3):125-139.
10. King CH, Dickman K, Tisch DJ: **Reassessment of the cost of chronic helmintic infection: a meta-analysis of disability-related outcomes in endemic schistosomiasis**. *Lancet* 2005, **365**(9470):1561-1569.
11. Crompton DW: **How much human helminthiasis is there in the world?** *J Parasitol* 1999, **85**(3):397-403.
12. Brown LA, Jones AK, Buckingham SD, Mee CJ, Sattelle DB: **Contributions from Caenorhabditis elegans functional genetics to antiparasitic drug target identification and validation: nicotinic acetylcholine receptors, a case study**. *Int J Parasitol* 2006, **36**(6):617-624.
13. Gasser RB, Cottee P, Nisbet AJ, Ruttkowski B, Ranganathan S, Joachim A: **Oesophagostomum dentatum: potential as a model for genomic studies of strongylid nematodes, with biotechnological prospects**. *Biotechnol Adv* 2007, **25**(3):281-293.
14. **Atlas of Pediatrics in the Tropics and Resource-Limited Settings**, 1 edn: American Academy of Pediatrics; 2009.
15. Sasaki T, Takagi M, Yaguchi T, Miyadoh S, Okada T, Koyama M: **A new anthelmintic cyclodepsipeptide, PF1022A**. *J Antibiot* 1992, **45**(5):692-697.
16. Ducray P, Gauvry N, Pautrat F, Goebel T, Fruechtel J, Desaules Y, Weber SS, Bouvier J, Wagner T, Froelich O *et al*: **Discovery of amino-acetonitrile derivatives, a new class of synthetic anthelmintic compounds**. *Bioorg Med Chem Lett* 2008, **18**(9):2935-2938.
17. WHO: **The selection and use of essential medicines**. *World Health Organ Tech Rep Ser* 2011(965):i-xiv, 1-249.
18. Prichard RK: **Anthelmintic resistance**. *Vet Parasitol* 1994, **54**(1-3):259-268.
19. Geerts S, Gryseels B: **Drug resistance in human helminths: current situation and lessons from livestock**. *Clin Microbiol Rev* 2000, **13**(2):207-222.
20. Prichard RK, Basanez MG, Boatin BA, McCarthy JS, Garcia HH, Yang GJ, Sripa B, Lustigman S: **A research agenda for helminth diseases of humans: intervention for control and elimination**. *PLoS Negl Trop Dis* 2012, **6**(4):e1549.
21. James CE, Hudson AL, Davey MW: **Drug resistance mechanisms in helminths: is it survival of the fittest?** *Trends Parasitol* 2009, **25**(7):328-335.
22. Lewis JA, Wu CH, Levine JH, Berg H: **Levamisole-resistant mutants of the nematode Caenorhabditis elegans appear to lack pharmacological acetylcholine receptors**. *Neuroscience* 1980, **5**(6):967-989.
23. Harrow ID, Gration KAF: **Mode of action of the anthelmintics morantel, pyrantel and levamisole on muscle cell membrane of the nematode Ascaris suum**. *Pestic Sci* 1985, **16**(6):662-672.
24. Martin RJ, Pennington AJ, Duittoz AH, Robertson S, Kusel JR: **The physiology and pharmacology of neuromuscular transmission in the nematode parasite, Ascaris suum**. *Parasitology* 1991, **102 Suppl**:S41-58.

25. Robertson SJ, Martin RJ: **Levamisole-activated single-channel currents from muscle of the nematode parasite Ascaris suum**. *Br J Pharmacol* 1993, **108**(1):170-178.
26. Kaplan RM: **Drug resistance in nematodes of veterinary importance: a status report**. *Trends Parasitol* 2004, **20**(10):477-481.
27. Lewis JA, Wu CH, Berg H, Levine JH: **The genetics of levamisole resistance in the nematode Caenorhabditis elegans**. *Genetics* 1980, **95**(4):905-928.
28. Fleming JT, Squire MD, Barnes TM, Tornoe C, Matsuda K, Ahnn J, Fire A, Sulston JE, Barnard EA, Sattelle DB *et al*: **Caenorhabditis elegans levamisole resistance genes lev-1, unc-29, and unc-38 encode functional nicotinic acetylcholine receptor subunits**. *J Neurosci* 1997, **17**(15):5843-5857.
29. Richmond JE, Jorgensen EM: **One GABA and two acetylcholine receptors function at the C. elegans neuromuscular junction**. *Nat Neurosci* 1999, **2**(9):791-797.
30. Culetto E, Baylis HA, Richmond JE, Jones AK, Fleming JT, Squire MD, Lewis JA, Sattelle DB: **The Caenorhabditis elegans unc-63 gene encodes a levamisole-sensitive nicotinic acetylcholine receptor alpha subunit**. *J Biol Chem* 2004, **279**(41):42476-42483.
31. Towers PR, Edwards B, Richmond JE, Sattelle DB: **The Caenorhabditis elegans lev-8 gene encodes a novel type of nicotinic acetylcholine receptor alpha subunit**. *J Neurochem* 2005, **93**(1):1-9.
32. Haugstetter J, Blicher T, Ellgaard L: **Identification and characterization of a novel thioredoxin-related transmembrane protein of the endoplasmic reticulum**. *J Biol Chem* 2005, **280**(9):8371-8380.
33. Almedom RB, Liewald JF, Hernando G, Schultheis C, Rayes D, Pan J, Schedletzky T, Hutter H, Bouzat C, Gottschalk A: **An ER-resident membrane protein complex regulates nicotinic acetylcholine receptor subunit composition at the synapse**. *Embo J* 2009, **28**(17):2636-2649.
34. Eimer S, Gottschalk A, Hengartner M, Horvitz HR, Richmond J, Schafer WR, Bessereau JL: **Regulation of nicotinic receptor trafficking by the transmembrane Golgi protein UNC-50**. *Embo J* 2007, **26**(20):4313-4323.
35. Halevi S, McKay J, Palfreyman M, Yassin L, Eshel M, Jorgensen E, Treinin M: **The C. elegans ric-3 gene is required for maturation of nicotinic acetylcholine receptors**. *Embo J* 2002, **21**(5):1012-1020.
36. Gottschalk A, Almedom RB, Schedletzky T, Anderson SD, Yates JR, 3rd, Schafer WR: **Identification and characterization of novel nicotinic receptor-associated proteins in Caenorhabditis elegans**. *Embo J* 2005, **24**(14):2566-2578.
37. Gally C, Eimer S, Richmond JE, Bessereau JL: **A transmembrane protein required for acetylcholine receptor clustering in Caenorhabditis elegans**. *Nature* 2004, **431**(7008):578-582.
38. Gendrel M: **[A new mode of nicotinic receptor clustering via a secreted CCP (complement control protein) containing protein]**. *M S-MED SCI* 2010, **26**(4):341-343.
39. Rapti G, Richmond J, Bessereau JL: **A single immunoglobulin-domain protein required for clustering acetylcholine receptors in C. elegans**. *Embo J* 2011, **30**(4):706-718.
40. Maryon EB, Coronado R, Anderson P: **unc-68 encodes a ryanodine receptor involved in regulating C. elegans body-wall muscle contraction**. *J Cell Biol* 1996, **134**(4):885-893.
41. Martin RJ, Robertson AP, Buxton SK, Beech RN, Charvet CL, Neveu C: **Levamisole receptors: a second awakening**. *Trends Parasitol* 2012, **28**(7):289-296.
42. Kwa MS, Veenstra JG, Roos MH: **Molecular characterisation of beta-tubulin genes present in benzimidazole-resistant populations of Haemonchus contortus**. *Mol Biochem Parasit* 1993, **60**(1):133-143.

43. Silvestre A, Cabaret J: **Mutation in position 167 of isotype 1 beta-tubulin gene of Trichostrongylid nematodes: role in benzimidazole resistance?** *Mol Biochem Parasit* 2002, **120**(2):297-300.
44. Ghisi M, Kaminsky R, Maser P: **Phenotyping and genotyping of Haemonchus contortus isolates reveals a new putative candidate mutation for benzimidazole resistance in nematodes**. *Vet Parasitol* 2007, **144**(3-4):313-320.
45. Njue AI, Hayashi J, Kinne L, Feng XP, Prichard RK: **Mutations in the extracellular domains of glutamate-gated chloride channel alpha3 and beta subunits from ivermectin-resistant Cooperia oncophora affect agonist sensitivity**. *J Neurochem* 2004, **89**(5):1137-1147.
46. Dent JA, Smith MM, Vassilatis DK, Avery L: **The genetics of ivermectin resistance in Caenorhabditis elegans**. *Proc Natl Acad Sci U S A* 2000, **97**(6):2674-2679.
47. Kohler P: **The biochemical basis of anthelmintic action and resistance**. *Int J Parasitol* 2001, **31**(4):336-345.
48. Boulin T, Fauvin A, Charvet CL, Cortet J, Cabaret J, Bessereau JL, Neveu C: **Functional reconstitution of Haemonchus contortus acetylcholine receptors in Xenopus oocytes provides mechanistic insights into levamisole resistance**. *Br J Pharmacol* 2011, **164**(5):1421-1432.
49. Robertson AP, Bjorn HE, Martin RJ: **Resistance to levamisole resolved at the single-channel level**. *FASEB J* 1999, **13**(6):749-760.
50. Sangster NC, Davis CW, Collins GH: **Effects of cholinergic drugs on longitudinal contraction in levamisole-susceptible and -resistant Haemonchus contortus**. *Int J Parasitol* 1991, **21**(6):689-695.
51. Varady M, Bjorn H, Craven J, Nansen P: **In vitro characterization of lines of Oesophagostomum dentatum selected or not selected for resistance to pyrantel, levamisole and ivermectin**. *Int J of Parasitol* 1997, **27**(1):77-81.
52. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M: **The transcriptional landscape of the yeast genome defined by RNA sequencing**. *Science* 2008, **320**(5881):1344-1349.
53. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q *et al*: **Full-length transcriptome assembly from RNA-Seq data without a reference genome**. *Nat Biotechnol* 2011, **29**(7):644-652.
54. Birol I, Jackman SD, Nielsen CB, Qian JQ, Varhol R, Stazyk G, Morin RD, Zhao Y, Hirst M, Schein JE *et al*: **De novo transcriptome assembly with ABySS**. *Bioinformatics* 2009, **25**(21):2872-2877.
55. Schulz MH, Zerbino DR, Vingron M, Birney E: **Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels**. *Bioinformatics* 2012, **28**(8):1086-1092.
56. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: **Mapping and quantifying mammalian transcriptomes by RNA-Seq**. *Nat Methods* 2008, **5**(7):621-628.
57. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M *et al*: **The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data**. *Genome Res* 2010, **20**(9):1297-1303.
58. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M *et al*: **A framework for variation discovery and genotyping using next-generation DNA sequencing data**. *Nat Genet* 2011, **43**(5):491-498.
59. Cantacessi C, Jex AR, Hall RS, Young ND, Campbell BE, Joachim A, Nolan MJ, Abubucker S, Sternberg PW, Ranganathan S *et al*: **A practical, bioinformatic workflow system for**

large data sets generated by next-generation sequencing. *Nucleic Acids Res* 2010, **38**(17):e171.

60. Mizrachi E, Hefer CA, Ranik M, Joubert F, Myburg AA: **De novo assembled expressed gene catalog of a fast-growing Eucalyptus tree produced by Illumina mRNA-Seq**. *BMC Genomics* 2010, **11**:681.
61. Tsai IJ, Otto TD, Berriman M: **Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps**. *Genome Biol* 2010, **11**(4):R41.

CHAPTER 3. COMPUTATIONAL CLONING OF ANTHELMINTHIC TARGET GENES IN THE NON-MODEL NEMATODE PARASITE *OESOPHAGOSTOMUM DENTATUM*

A paper to be submitted to the Journal of Parasitology

Nathan M. Romine, Richard J. Martin[*], Jeffrey K. Beetham[§]

Abstract

Sequence identification is a requirement for translating work done in *C. elegans* to non-model parasitic nematodes. This identification is currently done through imprecise degenerate PCR. Second generation sequencing offers a cheaper method of identifying these sequences. We present here an improved method for identifying full length sequences from SGS data using de novo assembly, refinement through resampling, and assignment through nearest –neighbor comparison. The utility of the algorithm is demonstrated by comparison to non-resampled assemblies. Resampling identified full coding sequences for 9 genes associated with anthelminthic resistance compared to just 2 full coding sequences from non-resampled assemblies. Two additional resampled sequences improved on the non-resampled assemblies, although full coding sequences were not determined. This methodology simplifies gene identification and translation into non-model organisms and has potential utility for the large scale identification/sequence-determination of target genes from any non-genetic-model (i.e. having an unknown genome) nematode.

Helminth infections of humans and domestic animals result in tremendous economic and social burdens throughout the world (Brooker, 2010). Treatment typically usually involves administration of an anthelmintic from one of the 3 main drug classes: the benzamidazoles, macrolytic lactones, and nicotinic agonists. Of particular concern are reports over the past 20+ years of resistance to members of each drug class (Geerts and Gryseels, 2000; James, Hudson et al., 2009).

Department of Veterinary Pathology, College of Veterinary Medicine, Iowa State University, Ames, Iowa 50011.
*Department of Biomedical Sciences, College of Veterinary Medicine, Iowa State University, Ames, Iowa 50011.
[§] Corresponding author

Studies of helminth resistance to these drugs, many performed in the free-living nematode *Caenorhabditis elegans*, have identified a number of genes that code for proteins which are drug targets of benzamidazoles (*ben-1*), macrolytic lactones (*avr-14, avr-15, glc-1, glc-2, glc-3, glc-4*), and the nicotinic agonists levamisole (*lev-1, lev-8, unc-29, unc-38, unc-63*) (reviewed in (Martin and Robertson, 2007; James, Hudson et al., 2009)). Many of these studies utilized *C. elegans* for a variety of reasons initially (e.g. cell simplicity, short life cycle, non-parasitic, favorable life-cycle traits towards genetic studies), and more recently for reasons that the sequence of its genome (and most of its genes) was known and that genetic tools were developed with which to finely manipulate gene expression. While it is also important to extend studies to parasitic nematodes, efforts at the molecular level are frequently hindered by the general lack of gene information for the parasitic nematodes. Consequently, a time-consuming process that is typically used to identify gene sequences in nematodes for which genomic data is unavailable involves (i) searching for the candidate gene from among known genes of genetically close organisms, (ii) aligning those gene sequences to determine regions of high similarity, (iii) designing DNA primers likely to align to the similar regions, and finally, (iv) amplifying the target using PCR and a single pair of primers, cloning and sequencing the PCR product.

A modification of this approach, one having the potential to reduce the time and expense required to identify single or multiple gene sequences in nematodes, is to use RNASeq data to build the gene sequences in silico, i.e. by computational methods. RNA-Seq is a second generation sequencing (SGS) technique that produces large amounts of sequence data for relatively little cost. Although RNA-Seq sequences are very short (50-150 bases long), when applied to an organism having a known genome RNA-Seq sequence data can be computationally analyzed using software packages such as the commonly used Velvet package (Zerbino and Birney, 2008) to produce in silico larger sequences (contigs) corresponding to entire gene sequences. Unfortunately, when the same

process is applied to organisms for which little genome information is known, the output is more typically comprised of short contigs and contigs of lower quality (Martin and Wang, 2011).

This report describes a method that builds upon the existing Velvet analysis software to allow greatly improved determination of gene sequences from *Oesophagostomum dentatum*, a nematode for which the genome is unknown. This method was used to determine the sequences of 12 *O. dentatum* genes that, in various nematode species, have been associated with resistance to several classes of anthelminthic drugs. Establishing these sequences will facilitate downstream physiological and molecular studies. In addition, the methodology has general utility in gene sequence determination for any organism that lack genomic information.

Materials and Methods

Parasites

Adult *O. dentatum* worms were harvested from a pig previously infected with a levamisole resistant isolate (Varady, Bjorn et al., 1997) as described (Robertson, Bjorn et al., 1999). Worms were counted, and separated by sex, by microscopic examination. To remove foreign contaminants, the separated (male versus female) worms were washed at least 3x in a pH 7.5 maintenance solution consisting of (mM): NaCl (150), KCl (2.7), $CaCl_2$ (2), $MgCl_2$ (0.3), PIPES (10), NaOH (13), glucose (11), $NaHCO_3$ (12), penicillin 0.06 g L^{-1}, streptomycin 0.1 g L^{-1}. Worms were settled by gravity before removing each wash, then blotted dry and transferred to a 1.5 ml microfuge tube, weighed by difference, and processed for extraction of RNA.

RNA isolation

Parasite samples resuspended in 1.0 ml TRI reagent (Molecular Research Center, Cincinnati, Ohio) were ground by mortar and pestle under liquid nitrogen then brought to a total volume of 2 to 3 ml TRI reagent. Total RNA was extracted from the TRI reagent according to the manufacturer's instructions, including an additional centrifugation step for clearing insoluble material. Extracted RNA was treated with DNase I (New England BioLabs, Ipswich, Massachusetts) (10 min at 37° C, 10

min at 75° C), then re-extracted with TRI reagent and resuspended in diethylpyrocarbonate-treated water. RNA concentration, purity, and quality (RNA Integrity Number) were assessed on a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California).

mRNA-Seq

The building of indexed, non-normalized, paired-end mRNA-Seq libraries, and subsequent 75-cycle pyrosequencing on an Illumina GAIIx platform, were performed as a service by the DNA Facility (Office of Biotechnology, Iowa State University) using 5 µg total RNA (per sample). Male and female libraries were duplexed in a single sequencing lane.

Genomics and bioinformatics

Assembly: Velvet version 1.1.06 (Zerbino and Birney, 2008) was used for contig assembly.

Similarity searching: BLAST algorithms (Altschul, Gish et al., 1990) were used to compare contigs with sequences available in public databases including the National Center for Biotechnology Information (NCBI) to identify homologues from other nematodes, i.e. sequences returning BLAST expect values $\leq 1E^{-10}$.

Read mapping: 64 bit Bowtie (Langmead, Trapnell et al., 2009) version 0.12.7 was used to map reads for contig building.

Pairwise comparison: The Needle algorithm (Needleman and Wunsch, 1970) was used for pairwise comparison.

Custom codes: Java 1: Java-script code reads BLAST output and collects contigs that pass identity thresholds from a contig source file (APPENDIX A). The length and identity of BLAST HSPs collected are adjustable parameters in the source code and should be adjusted for the degree of relatedness between the organisms compared.

Java-2: Java-script code that reads Bowtie alignment file and collects mapped and orphan reads from read source file (APPENDIX B).

Results

Generation of mRNA-Seq read libraries

Separate mRNA-Seq libraries were constructed from 5 µg of high quality total RNA (RNA Integrity Number $\geq$ 7.3) isolated from 58 adult male and 141 adult female levamisole resistant worms. Given the absence of a sequenced genome of *O. dentatum* for use as a scaffold during subsequent assembly and analysis, paired-end sequencing was performed to facilitate contig-building steps. Similar numbers of 75-cycle paired reads were obtained for both libraries, (2.144×10^7 for male, 2.159×10^7 for female) for a combined total of more than 40 million reads. Reads were trimmed of read tags (added during library building to allow multiplexing during the sequencing run) and deposited at the NCBI sequence read archive: male library (SRR393668) and female library (SRR393669).

Velvet assembly

Once mRNA-Seq read libraries are constructed, the first 2 steps towards identifying specific genes and their sequences in silico from mRNA-Seq data are to assemble the mRNA-Seq reads into contigs (i.e. contiguous sequences) and then to identify those contigs that have high similarity to the specific gene targets (Fig. 3.1, steps 1-2). The SGS assembler Velvet was used to assemble the 40 million reads from the combined male and female libraries into contigs. Combining the 2 libraries increased the potential number of reads that were identified for each target sequence while maintaining compliance with NCBI guidelines requiring submitted sequences to be derived from a single strain. As noted previously, sequencing the libraries as paired end reads, wherein sequencing of a single molecule results in 2 reads (one for each end of the molecule) allowed Velvet to assemble using its paired-read mode within which an individual read is used to assemble a particular contig only when the corresponding read from the other end of the same molecule is also used in the same contig. One early step within Velvet results in individual reads being hashed (indexed) into overlapping sequences (k-mers) of some length “k” that has influence over downstream analyses. The

k-mer sequences are obtained as if a sliding window of length "k" is moved along a given read. Thus a 10 base DNA read sequence "AGGACTTAGA", when hashed with a "k-mer" length of 7, would have a first hash sequence of "AGGACTT", a second of "GGACTTA", a third of "GACTTAG", and a fourth (final) of "ACTTAGA". In general, the k-mer setting affects length and quality of the contigs built during an assembly, with longer k-mers generally corresponding to greater specificity (i.e. that a detected alignment is actually correct) but lower sensitivity to detect alignments (Zerbino and Birney, 2008). Because optimal k-mer length cannot be known a priori, multiple assemblies were performed using k-mers ranging from length 17 to 49.

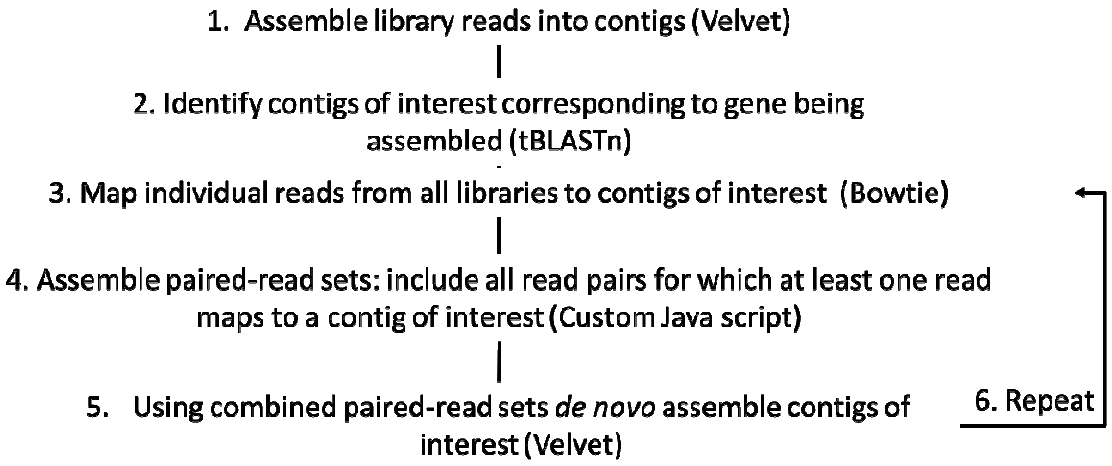

FIGURE 3.I - Algorithm for in silico determination of gene sequences. [1] Combined read sets were assembled into contigs using Velvet with a K-mer of 31. [2] BLAST searches for anthelminthic target genes were carried out with a high cutoff (expect value = 1E-10) to identify contigs representing target genes. [3] Reads from all data sets were individually mapped using Bowtie (e 150) to contigs identified by database searches. [4] Paired reads where at least one read mapped to a contig were gathered using a custom Java script. [5] De novo assembly of sequences was performed using Velvet. [6] Mapping, pair gathering, and de novo assembly were repeated until the iteration resulted in no additional reads being mapped to the contig of interest.

The 12 anthelminthic resistance associated genes of *C. elegans* for which we sought to determine the homologous genes and sequences of *O. dentatum* are shown in Table 3.I; the drug groups to which they associate are shown in Table 3.II. To identify contigs exhibiting high similarity to the *C. elegans* target sequences we first built separate databases, each comprised of the set of contigs produced by a single k-mer assembly, and then queried each database using the protein

sequences of the 12 anthelminthic resistance associated genes. Specifically, databases were first created as BLAST databases to facilitate subsequent searches using BLAST programs. Then, contigs with high similarity to the target proteins were identified by querying the database with the tblastn program, a program in which a protein sequence queries a nucleotide database that has been translated in all reading frames and in both directions. A stringent "E-value" threshold of 1×10^{-10} was used within the tblastn search in order to limit false positive identifications; the E-value corresponds to the number of times a match of the same quality would be found among the database by chance. A custom java-script (APPENDIX A) was used to collect those contigs whose E-values were $\leq 1 \times 10^{-10}$ and, for further stringency, to retain only those contigs whose region of alignment with the target protein exhibited $\geq 60\%$ identity across $\geq 50\%$ of the contig length. Some contigs that made it past these filters exhibited high similarity to more than one target gene (data not shown), as would be expected since some of the target genes are paralogous (deriving from an historical gene duplication event) and thus are expected to share high similarity. To unambiguously assign the filtered contigs to single target genes, contigs were used to query the *C. elegans* genome database (using the BLAST program blastx, which searches a protein database using a translated nucleotide query sequence), and then gene identity was assigned based upon the match exhibiting the lowest E-value.

For each gene of interest, Table 3.I summarizes the BLAST profile for each k-mer assembly. BLAST determines the region (called high scoring pair, or HSP in Tables 3.I and 3.II) within a contig that aligns with high similarity to a target sequence. While it is possible for a single contig to yield multiple HSPs, as occurs when regions of high similarity are separated by regions of below-threshold similarity, all contigs identified here yielded only a single HSP (i.e. aligned region) per target gene. Table 3.I demonstrates that, for a given length k-mer, the quality of the assembly varies greatly, with (i) some k-mers producing many but short HSPs, (ii) other k-mers yielding few, but longer HSPs, and (iii) no single k-mer yielding the longest HSP (contig) for all target genes. Table 3.I also shows that

Table 3.I - Target identification from 19 increasing k-mer assemblies

Max HSP length corresponds to the longest HSP identified among all k-mers used. # HSP is the number of HSPs that were positively identified for each gene at the k-mer indicated. Avg HSP length is the average length of those HSPs. HSP range is the minimum and maximum length of those HSPs identified for each gene/k-mer.

Target Genes		Initial library assemblies																	
		Best	17 mer			19 mer			21 mer			23 mer			25 mer			27 mer		
GENE ID	CDS len	HSP length	# HSPs	avg HSP length	HSP range	# HSPs	avg HSP length	HSP range	# HSPs	avg HSP length	HSP range	# HSPs	avg HSP length	HSP range	# HSPs	avg HSP length	HSP range	# HSPs	avg HSP length	HSP range
lev-1	1419	1401				2	205	132-279	3	148	99-219	3	238	174-294	5	238	96-525	4	237	114-381
lev-8	1596	255							2	147	126-168	1	255	**255-255**	1	108	108-108			
unc-29	1482	480				4	153	138-168	12	144	105-168	7	259	**120-480**	8	248	132-426	11	202	96-402
unc-38	1524	1473	1	111	111-111	7	175	111-294	6	180	111-294	9	184	90-375	9	283	84-1098	8	339	87-1380
unc-63	1524	576				4	174	105-207	3	272	**105-576**	5	247	96-510	5	263	105-510	5	235	105-510
avr-14	1251	390				4	124	66-261	1	90	90-90	2	112	102-123	2	114	105-123	4	111	105-123
avr-15	1437	291	2	106	93-120	2	186	162-210	4	163	123-210	7	154	123-192	6	140	105-210	3	185	123-240
ben-1	1335	453							2	163	108-219	1	111	111-111	1	114	114-114	3	102	90-117
glc-1	1305	318							1	123	123-123	1	108	108-108	1	318	**318-318**	1	237	237-237
glc-2	1305	939	1	183	183-183	1	285	285-285	4	198	147-318	1	879	879-879	1	939	**939-939**	3	375	207-582
glc-3	1455	1383				1	168	168-168	3	146	87-207	2	115	111-120	2	102	93-111	4	117	99-144
glc-4	1503	1428				4	143	108-225	6	192	111-279	2	702	249-1155	2	702	249-1155	2	702	249-1155

Target Genes		Best	29 mer			31 mer			33 mer			35 mer			37 mer			39 mer		
GENE ID	CDS len	HSP length	# HSPs	avg HSP length	HSP range	# HSPs	avg HSP length	HSP range	# HSPs	avg HSP length	HSP range	# HSPs	avg HSP length	HSP range	# HSPs	avg HSP length	HSP range	# HSPs	avg HSP length	HSP range
lev-1	1419	1401	5	219	93-420	11	150	99-330	9	181	114-255	8	207	102-558	6	334	138-657	3	564	339-807
lev-8	1596	255																		
unc-29	1494	480	8	223	120-465	9	171	96-378	8	180	96-396	2	172	162-183	4	173	144-207			
unc-38	1524	1473	7	493	**90-1473**	7	416	**84-1473**	6	472	**108-1473**	4	708	**261-1473**	3	890	**360-1473**	1	1473	**1473-1473**
unc-63	1524	576	3	227	105-297	4	285	105-414	3	275	105-414	1	414	414-414	2	277	141-414			
avr-14	1251	390	3	123	111-135	5	145	111-207	6	169	123-270	5	202	123-321	5	240	123-327	1	390	**390-390**
avr-15	1437	291	5	156	90-243	7	153	**96-291**	4	132	93-213	4	138	96-243	4	108	108-111	2	111	111-111
ben-1	1335	453	2	97	93-102	4	99	90-120	3	99	96-102	3	109	87-135	5	109	93-129	7	100	75-120
glc-1	1386	318	1	237	237-237				2	141	141-141									
glc-2	1305	939	3	261	147-393	2	439	135-744	2	475	207-744	2	387	171-603	2	336	333-339			
glc-3	1455	1383	3	112	111-114	7	169	93-330	7	120	90-168	5	145	117-213	6	153	111-189	3	618	**84-1383**
glc-4	1503	1428	2	702	249-1155	2	702	249-1155	2	702	249-1155	2	702	249-1155	1	972	972-972	1	1428	**1428-1428**

Target Genes		Best	41 mer			43 mer			45 mer			47 mer			49 mer		
GENE ID	CDS len	HSP length	# HSPs	avg HSP length	HSP range	# HSPs	avg HSP length	HSP range	# HSPs	avg HSP length	HSP range	# HSPs	avg HSP length	HSP range	# HSPs	avg HSP length	HSP range
lev-1	1419	1401	1	1392	1392-1392	1	1401	**1401-1401**	2	619	345-894	3	245	219-270	2	256	162-351
lev-8	1596	255															
unc-29	1494	480															
unc-38	1524	1473	1	1473	**1473-1473**	2	868	**264-1473**	1	1473	**1473-1473**						
unc-63	1524	576															
avr-14	1251	390	1	219	219-219												
avr-15	1437	291				1	120	120-120	1	174	174-174	1	174	174-174			
ben-1	1335	453	10	138	90-234	16	115	84-234	25	114	87-348	15	160	**90-453**	15	150	93-312
glc-1	1386	318															
glc-2	1305	939															
glc-3	1455	1383				1	378	378-378									
glc-4	1503	1428	1	1428	**1428-1428**	1	972	972-972	1	972	972-972	1	1428	**1428-1428**			

the Velvet assembly produced full or nearly full length sequences for 4 genes (*lev-1*, *unc-38*, *glc-3*, and *glc-4*), and relatively short contigs for the other 8 targets.

Resampling to optimize sequence determination

To improve the overall quality of the contigs (i.e. increase contig lengths by extension and by gap filling) additional computational steps of assembly were developed (Fig. 3.1, steps 3-5). In outline, this involved identifying all RNA-Seq library reads that matched a particular contig and then using Velvet to reassemble those reads into contigs; this process was repeated until the output contigs exhibited no relative improvement. In detail, high-identity contigs that were identified in the initial Velvet assembly (at k-mer length 31 for all gene targets excepting *lev*-8, *glc-1*, and *glc*-3,which used k-mer lengths of 23, 25, and 39 respectively) were used to identify all library reads that mapped to those contigs; this process utilized the mapping program Bowtie set to increase the sensitivity of read identification by using a low quality threshold value (150) and by running it in unpaired-read mode. The unpaired-read mapping mode allows inclusion of those reads that did not contribute to the initial contig (Table 3.I) because their pair failed to map to that initial contig. Consequently, the collected reads included the paired reads along with a number of single reads for which their pair did not map; a custom Java-script (Fig. 3.1, Step 4: APPENDIX B) collected into a single bin all of the reads that mapped as well as reads that did not map but whose pair did map. The Velvet program was then used to assemble contigs from the collected paired-end reads combined from both libraries (Fig. 3.1, Step 5) using multiple length k-mers and coverage cutoffs to identify assembly conditions that produced a maximum contig length. Bowtie mapping, read set collection and assembly were repeated (Fig. 3.1, Step 6) until a maximum contig length and maximum coverage (relative to the target genes) were achieved.

The dramatic effectiveness of the resampling is evidenced by the comparison of resampled contigs, initial contigs (best HSP), and *C. elegans* target sequence shown in Table 3.II. Whereas the initial velvet assembly yielded 2 sequences corresponding to 100% of the target gene coding

sequences (*unc-38*, *glc-4*), resampling yielded the full coding sequence of an additional 7 target sequences including a second isotype of *ben-1*, and yielded large improvements to 2 gene sequences (*unc-29*, *unc-63*).

Table 3.II – Target gene identification and comparison

Accession numbers are given for DNA sequences assembled from read libraries and sequences used to assemble. Best initial HSP "Length" and "C length" are the length of the longest HSP from the raw read assemblies and the length of the contig from which that HSP derives. "% P" represents the percent of the comparator *C. elegans* protein that is encoded by the corresponding resampled contig. "% Identity" represents results from a pairwise alignment (performed using the Needle algorithm) of the comparator and of the protein deduced from the resampled contig. * sequences have been submitted to NCBI. These sequences are listed by gene ID in APPENDIX C.

Resampled contigs					Best initial HSP		C elegans sequence		Comparison results	
Gene ID	Accession #	Length	HSP Length	# AA	C length	Length	Accession #	# AA	% P	% Identity
Levamisole resistance genes (nAChR subunits)										
lev-1	submitted*	1612	1401	477	1580	1401	CAB03148	472	100.00	91.40
lev-8	submitted	269	255	89	269	255	CAB01685	531	16.76	85.39
unc-29	submitted	614	576	204	482	480	CAB02308	493	41.38	87.75
unc-38	submitted	1631	1455	507	1607	1473	CCD69819	507	100.00	72.25
unc-63	submitted	1770	1248	417.67	578	576	CCD66192	507	81.85	81.93
Avermectin, benzamidazole resistance genes										
avr-14	submitted	1485	1299	464	629	390	CCD61323	416	100.00	52.48
avr-15	submitted	1531	1242	447	560	291	CAB03329	478	100.00	44.29
ben-1_1	submitted	1436	1308	448	453	453	CAB00853	444	100.00	94.21
ben-1_2	submitted	1751	1317	448	453	453	CAB00853	444	100.00	95.09
glc-2	submitted	1389	1215	424	994	939	CCD62432	434	100.00	71.69
glc-3	submitted	2048	1383	531	1724	1383	CCD69051	484	100.00	57.87
glc-4	submitted	1628	1521	508	1628	1428	CCD65896	500	100.00	77.10

Interestingly, while the initial assembly yielded a contig that best-mapped by BLAST analysis to *glc-1* (Table 3.I), resampling yielded an extended contig that was unambiguously identified as *avr-15*. A closer examination of the contigs produced during the initial assembly that were identified as *glc-1* revealed that they exhibited almost-as-high similarity to *avr-15* (data not shown). The reason for the initial misidentification is that the region of similarity to *glc-1* is relatively small and is highly conserved with a region of *avr-15*, a paralog of *glc-1*.

The quality/accuracy of the in silico derived sequences can be inferred from the shared identity to target genes of *C. elegans* that is shown in Table 3.I, and, for the two target genes (*unc-63, unc-38)* whose *O. dentatum* mRNA sequences had previously been deposited in Genbank, by comparison of the "old" and "new" sequences. BLAST nucleotide comparison of *O. dentatum unc-*

63 mRNA sequence (accession HQ162136.1) to the corresponding 1770 nucleotide in silico sequence identified a single alignment comprised of 1768 bases (including 30 non-identical bases) and no gaps. A similar comparison of the *O. dentatum* unc-*38* mRNA sequence (accession GU256648.1, length 1681) to the corresponding 1631 nucleotide in silico sequence identified a single alignment comprised of 1603 bases (including 4 non-identical bases). These comparisons indicate the high quality of *unc-38* and the *unc-63* in silico sequences and, by inference, of the other in silico sequences reported here.

Discussion

The data shown in Results demonstrate the efficient and successful use of an iterative de novo assembly of RNASeq data to determine in silico the sequence of 11 *O. dentatum* genes that are implicated in anthelminthic resistance. Selection of these particular genes was based upon the results of a number of studies (reviewed in (Martin and Robertson, 2007; James, Hudson et al., 2009)). The iterative assembly produced full length coding sequences for 9 target genes, whereas the Velvet assembly yielded full length sequences for only a 2-gene subset of those 9 genes. A major utility of this process is that, as a computational process, it is scalable and should fit well in a variety of situations, whether the need is to identify sequences for 10 genes or for 100 genes of *O. dentatum* or of any other nematode lacking a known genome sequence. At the same time, once the initial assemblies have been made, the process is computationally efficient because it reduces the need for full library assemblies by assembling only that subset of library reads that map to the initial (or subsequent) high-identity contigs.

Other related computational processes have been described that are designed to yield different output than the full length coding sequences that are described here. For example, one group demonstrated that remapping of reads by identity to contigs within an initial assembly, and then reassembling contigs from those remapped reads, improved transcriptome assembly (Mizrachi, Hefer et al., 2010); the desired output from that work was production of a general gene ontology. Another group interested in improving genome assemblies showed varied levels of success in gap closure by

mapping paired-end reads and collecting pairs for which only one of the ends aligned to a contigs (Tsai, Otto et al., 2010).

As shown in Table 3.I, a number of k-mer lengths were used for initial library assemblies, demonstrating the dramatic effect of k-mer length on contigs. That said, as noted in RESULTS, k-mer length 31 was used to build the contigs used in all target-gene resampling excepting that for *lev-8* (for which k-mer 31 returned no contigs; see Table 3.I) and *glc-3* (for which k-mer 31 contigs failed to support generation of resampled contigs representing the full target sequence). This suggests there is little need to use the best or longest contig as input to the resampling process. As a further test of this concept we were able to successfully build *avr-15* from a single approximately 300 base initial contig instead of from the 7 contigs shown in Table 3.I. Thus, there is little need to conduct full library assemblies over a wide range of k-mers when attempting to derive in silico the sequences for a set of target genes; instead, one stops when a limited set of k-mers that have been run produce a quality contig for each target.

Of interest is the identification of a *lev-8*-like sequence. This identification is based upon the sequence being 85% identical to *C. elegans lev-8* at the deduced protein level (Table 3.II). By comparison, pairwise alignment of the *lev-8*-like sequence to the *O. dentatum* or *Haemonchus contortus acr-8* indicated only 67% identity at the deduced protein level. Although *lev-8* is found in *C. elegans* (*C. elegans* Sequencing Consortium, 1998) it has not been found in *H. contortus* (Boulin, Fauvin et al., 2011), a clade V nematode that is considered more closely related to *O. dentatum* than is *C. elegans* (Blaxter, De Ley et al., 1998). These data suggest a loss of *lev-8* has occurred in *H. contortus*. Because data reported here identified only a very short *lev-8* contig (<300bp), it remains to be determined whether or not *O. dentatum* contains a functional *lev-8*, as compared to a vestigial (partial) *lev-8* sequence. Functional *lev-8* nAChR subunit has been shown to confer sensitivity to levamisole (Towers, Edwards et al., 2005); identification of the complete gene in *O. dentatum* may

explain differences in the nAChR properties of closely related nematode species and potentially of differences in drug efficacy.

The identification of full coding sequences for these genes involved in anthelminthic resistance opens several areas for further examination. Primary among the goals of this study was the ability to identify full coding sequences to allow reconstitution and functional assay of these proteins in an in vitro system (similar to (Boulin, Fauvin et al., 2011)); these studies are underway. With data from additional isolates it will be possible to compare sequence and expression patterns between isolates. This could lead to identification of SNPs that produce functional resistance or of RNA expressions patterns that implicate specific protein levels in resistance. It is also possible to use this approach on a larger scale to facilitate gene ontology projects. More complete transcriptomic assembly would simplify gene identification and would particularly benefit organisms without genomes.

Acknowledgements

We thank Hui-Hsien Chou (Iowa State University) for assistance. Research funding was by The Hatch Act, State of Iowa and by the NIH grant: R01 AI047194 National Institute of Allergy and Infectious Diseases. The funding agencies had no role in the design, execution or publication of this study. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Allergy and Infectious Diseases.

Literature Cited

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. 1990. Basic local alignment search tool. Journal of Molecular Biology **215**: 403-410.

Blaxter, M. L., De Ley, P., Garey, J. R., Liu, L. X., Scheldeman, P., Vierstraete, A., Vanfleteren, J. R., Mackey, L. Y., Dorris, M., Frisse, L. M., et al. 1998. A molecular evolutionary framework for the phylum Nematoda. Nature **392**: 71-75.

Boulin, T., Fauvin, A., Charvet, C. L., Cortet, J., Cabaret, J., Bessereau, J. L. and Neveu, C. 2011. Functional reconstitution of Haemonchus contortus acetylcholine receptors in Xenopus oocytes provides mechanistic insights into levamisole resistance. British Journal of Pharmacology **164**: 1421-1432.

Brooker, S. 2010. Estimating the global distribution and disease burden of intestinal nematode infections: adding up the numbers--a review. International Journal for Parasitology **40**: 1137-1144.

C. elegans Sequencing Consortium 1998. Genome sequence of the nematode C. elegans: a platform for investigating biology. Science **282**: 2012-2018.

Geerts, S. and Gryseels, B. 2000. Drug resistance in human helminths: current situation and lessons from livestock. Clinical Microbiology Reviews **13**: 207-222.

James, C. E., Hudson, A. L. and Davey, M. W. 2009. Drug resistance mechanisms in helminths: is it survival of the fittest? Trends in Parasitology **25**: 328-335.

Langmead, B., Trapnell, C., Pop, M. and Salzberg, S. L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology **10**: R25.

Martin, J. A. and Wang, Z. 2011. Next-generation transcriptome assembly. Nature Reviews Genetics **12**: 671-682.

Martin, R. J. and Robertson, A. P. 2007. Mode of action of levamisole and pyrantel, anthelmintic resistance, E153 and Q57. Parasitology **134**: 1093-1104.

Mizrachi, E., Hefer, C. A., Ranik, M., Joubert, F. and Myburg, A. A. 2010. De novo assembled expressed gene catalog of a fast-growing Eucalyptus tree produced by Illumina mRNA-Seq. BMC Genomics **11**: 681.

Needleman, S. B. and Wunsch, C. D. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. Journal of molecular biology **48**: 443-453.

Robertson, A. P., Bjorn, H. E. and Martin, R. J. 1999. Resistance to levamisole resolved at the single-channel level. FASEB Journal **13**: 749-760.

Towers, P. R., Edwards, B., Richmond, J. E. and Sattelle, D. B. 2005. The Caenorhabditis elegans lev-8 gene encodes a novel type of nicotinic acetylcholine receptor alpha subunit. Journal of Neurochemistry **93**: 1-9.

Tsai, I. J., Otto, T. D. and Berriman, M. 2010. Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. Genome Biology **11**: R41.

Varady, M., Bjorn, H., Craven, J. and Nansen, P. 1997. In vitro characterization of lines of Oesophagostomum dentatum selected or not selected for resistance to pyrantel, levamisole and ivermectin. International Journal for Parasitology **27**: 77-81.

Zerbino, D. R. and Birney, E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Research **18**: 821-829.

CHAPTER 4. TRANSCRIPTOMIC EVALUATION OF THE NICOTINIC ACETYLCHOLINE RECEPTOR PATHWAY IN LEVAMISOLE-RESISTANT AND -SENSITIVE *OESOPHAGOSTOMUM DENTATUM*

A paper to be submitted to BMC Genetics

Nathan M. Romine[1], Jack M. Gallup[1], Richard J. Martin[2], Jeffrey K. Beetham[1§]

Abstract

Background

Nematode parasites infect humans and animals. The number of anthelminthic drugs available for treatment is limited, and nematode populations have developed resistance to many of them including the nicotinic agonist levamisole. It has been demonstrated that changes in the nicotinic acetylcholine receptor (nAChR) signaling pathway are associated with resistance to levamisole in parasitic nematodes. Furthermore, mutations in genes of the nAChR signaling pathway have been shown in the free-living nematode *Caenorhabditis elegans* to produce resistance to levamisole.

Results

This study of the parasitic nematode *Oesophagostomum dentatum* compared levamisole-sensitive and –resistant isolates using mRNA-Seq data to identify and characterize mRNA sequences corresponding to genes of the nAChR pathway. The full length mRNA sequences of 16 genes of the nAChR pathway were identified. Expression levels characterized from the transcriptomic data showed two-fold or greater reduction in expression in the resistant isolate for the genes *lev-10*, *lev-11*, *nra-1*, *ric-3*, *unc-22*, *unc-38*, *unc-63*, and *unc-68*$_{ry}$. SNP analysis showed a high rate of polymorphism within the two isolates; however, only 72/2788 of the identified SNPs caused amino acid changes.

[1] Veterinary Pathology, Iowa State University, Ames, IA

[2] Biomedical Sciences, Iowa State University, Ames, IA

[§] Corresponding author

Conclusions

Reduced levels of receptors containing the subunits UNC-38 and/or UNC-63, and decreased expression of LEV-10 and NRA-1, required for nAChR aggregation, LEV-11, the tropomysin (muscle) gene, UNC-22, which encodes twitchin and is required for the muscle contraction-relaxation cycle, and UNC-68$_{ry}$, a calcium release protein, associate with levamisole resistance; further analysis is required to determine whether this relationship is causal.

Keywords

Levamisole resistance, *Oesophagostomum dentatum*, nicotinic acetylcholine receptor

Introduction

Intestinal nematode parasites are of concern for both human and animal health, infecting more than 1 billion people worldwide in addition to animals of all species, causing significant morbidity and mortality [1]. Treatment of nematode infections is largely afforded by drug therapy using anthelminthics that fall within one of 3 major classes of compounds: benzamidazoles, macrolytic lactones, or nicotinic agonists. Specific anthelminthics used for treatment of human nematode infections include those on the WHO Model List of essential medicines: albendazole, levamisole, mebendazole, niclosamide, praziquantel, and pyrantel [2]. Resistance to anthelminthics has been reported in agricultural use to all 3 major drug classes although not the most recently introduced novel classes of anthelmintic: cyclooctadepsipeptides and amino acetonitrile derivatives [3]. Data also show increasing treatment failures, an indicator of drug resistance, in human parasitic worm infections [4, 5].

Levamisole, a nicotinic agonist, has been critical in allowing economically feasible treatment of animals throughout the world given its broad spectrum of activity and effectiveness against many larval stages. Exposure of adult helminths to levamisole results in spastic paralysis, due to somatic muscle contraction, that leads to expulsion from the gut. Resistance to levamisole has become an increasing problem since resistance was first reported in agricultural use in 1979 [6]. Investigations of

levamisole resistance using the free-living model nematode, *C. elegans*, have identified a number of phenotypes and genes coding for subcellular molecular pathways that associate with laboratory-induced resistance (reviewed in Martin, et al [7]). As would be expected based upon a mode of action that involves muscle paralysis, a commonly identified resistance associated pathway is that of the *C. elegans* nicotinic acetylcholine receptor (nAChR), a receptor pathway that controls the body wall muscles via the ion-channel activity of the receptor. The nAChR pathway is also implicated in resistance in a number of other Clade V nematodes including *Haemonchus contortus*, *Teledorsagia circumcincta* and *Trychostrongylus colubriformis* [8]. The nAChR is a heteropentameric ion channel comprised of subunit proteins encoded by over 25 different genes grouped within 5 gene families [7].

Electrophysiological studies of levamisole resistance in the Clade V swine parasitic nematode *O. dentatum* have shown that as levamisole concentrations were increased, levamisole activated channels in a levamisole-resistant isolate exhibited lowered responsiveness to levamisole as compared to a levamisole-sensitive isolate [9]. This change in the receptor response could result from changes to the levamisole receptor subunit population or subunit sequences or changes in subunit post-translational modifications.

The electrophysiological studies have provided extremely useful information towards increasing our understanding of anthelminthic resistance. To extend these studies to the level of gene expression assessment, the present study using mRNA-Seq sought to identify the sequences of genes of the nAChR pathway, and to characterize their mRNA expression and sequence changes in levamisole-resistant (LEVR) and -sensitive (SENS) isolates of *O. dentatum*.

Results

mRNA-Seq

Paired-end mRNA-Seq libraries were constructed from 5 ug high quality (RNA Integrity Number $\geq$ 7.2) total RNA of adult *O. dentatum* populations (SENS males and females) and then sequenced. Paired-end sequencing was performed to facilitate assembly and analysis. Similar

numbers of paired reads were obtained for both libraries, ranging from 2.012 X 10^7 to 2.108 x 10^7. Read libraries were deposited at the NCBI sequence read archive: SENS male (SRR393670), SENS female (SRR393671).

Equivalent read libraries for a levamisole resistant isolate were previously described [10]; these libraries are also available at the NCBI sequence read archive: LEVR male (SRR393668), LEVR female (SRR393669).

Sequence identification

The mRNA-Seq data analysis targeted the 26 genes of the nAChR pathway that are listed in Table 4.1, a pathway previously implicated in levamisole resistance in *C. elegans* and *O. dentatum* (see Introduction). Of the 26 genes, the coding sequence was unknown for 19 genes and only partially determined for another (*unc-29*$_{alt}$). The coding sequences for those 19 genes were determined in silico by assembling the library sequence reads *de novo* using the Velvet assembler and by a method of resampling as described previously [10]. In brief, high identity matches between the comparator genes (Table 4.1) and initial contigs assembled with Velvet , matches with an expect (E) value lower than 1 x 10^{-10} that also exhibited ≥ 60% identity over ≥ 50% of the contig length, were identified as target contigs representing the target genes. All read-pairs for which at least one read mapped by Bowtie to the target contigs were then placed into a bin and re-assembled using Velvet; this process was reiterated until maximum length and quality were obtained, or until further iteration failed to improve the contig. For *nra-1,* a lowering of the threshold to ≥50% identity over ≥ 50% of the contig length was necessary to identify contigs for successful reiterative assembly. Table 4.1 summarizes details on the comparator genes used to identify the homologous *O. dentatum* transcript sequences and the sequences assembled in silico.

The identities of the deduced protein sequences of O. dentatum genes as compared to that of the comparator genes were calculated using the Needle algorithm (Table 4.1). According to the classification scheme of Clade 5 nematodes [11], O. dentatum is more closely related to H. contortus

than to the various species of the Caenorhabditis genus. Similarly, the identity scores for 5 O. dentatum genes compared to their H. contortus and C. elegans homologs were higher, overall (ranging from 62.17 to 91.61, with mean/median = 79.16/84.54), than were the corresponding identities to the comparator genes of C. elegans species (44.85-77.98, mean/median = 64.29/70.45).

Table 4.1 - Sequences identified for nAChR-pathway genes

Accession numbers are given for sequences identified and for sequences used for comparison. Pairwise comparisons were performed using the Needle algorithm. % CDS calculates the number of amino acids in the comparator sequence represented in the identified sequence. % Identity is the % amino acid identity. Sequences marked with a star (*) were assembled from a single isolate for submission, but from both isolates to complete the coding sequence for analysis.

Assembled Contigs				Comparator			Comparison Results	
GENE ID	Accession #	LENGTH	# AA	SPECIES	Accession #	# AA	% P	% IDENTITY
Receptor Subunits								
acr-8	GAAS01000001	1792	538	O dentatum	JX429921	538	100.00	99.81
acr-16	GAAS01000002	1576	498	H contortus	ABW07339	494	100.00	84.54
acr-20	GAAS01000003	1785	561	C elegans	NP_001122627	564	100.00	61.70
acr-21	GAAS01000004	1410	456	C elegans	CAP33541	469	91.47	69.00
acr-25	GAAS01000005	846	290	C elegans	NP_001023570	544	46.88	73.13
lev-1	GAAR01000001	1534	477	H contortus	ADF56004	472	100.00	91.40
lev-8	GAAS01000006	269	89	C elegans	NP_509932	531	16.76	85.39
unc-29	n/a	889	297	O dentatum	JX429919	497	59.76	100.00
unc-29$_{alt}$	n/a	942	306	O dentatum	JX863910	261	117.24	99.62
unc-38	GAAR01000011	1744	507	O dentatum	GU256648	507	100.00	99.41
unc-63	GAAR01000012	2211	507	O dentatum	HQ162136	507	100.00	99.80
Non-Subunit Pathway Proteins								
lev-9	GAAR01000003	4179	1346	C remanei	XP_003118435	1381	100.00	57.97
lev-10	GAAR01000002	2745	902	C elegans	CAE82298	906	100.00	61.96
lev-11	GAAS01000007	1088	284	C elegans	NP_001021695	284	100.00	95.42
nra-1	GAAS01000008	2428	728	C elegans	CAD56248	622	100.00	38.70
nra-2	GAAR01000004	1717	544	C elegans	CAB04692	557	100.00	51.25
nra-4	GAAS01000009	3585	1104	C elegans	CCD62613	1121	100.00	47.93
oig-4	GAAR01000005	526	150	C elegans	NP_001022278	155	100.00	76.77
ric-3	GAAR01000006	1323	358	H contortus	ADV92281	365	100.00	62.17
soc-1	GAAR01000007	1674	473	C elegans	NP_504250	430	100.00	27.34
tax-6	GAAR01000008	2170	516	C elegans	CAB02719	535	100.00	84.51
tpa-1	GAAR01000009	2813	707	C elegans	NP_499860	704	100.00	88.26
unc-22	GAAR01000010	20890	6755	C elegans	NP_502273	6839	100.00	72.52
unc-50	GAAS01000010	1390	303	H contortus	ADV92279	298	100.00	86.23
unc-68$_{ry}$	GAAR01000013	15639	5075	C elegans	BAA08309	5071	100.00	80.60
unc-74	GAAS01000011	1559	445	H contortus	ADV92278	445	100.00	71.24

Table 4.2 – RNA-seq Expression profiles of nAChR-pathway genes

(A): Reads were mapped against the coding sequence of assembled sequences shown in Table 4.1. Normalized data are presented as FPKM, with the number of reads mapped shown in parentheses. (B): Values represent the ratio of the normalized expression data shown in A for isolates/libraries denoted by column headings. Numbers in bold represent expression differences greater than two-fold.

(A)	LEVR-MALE	SENS-MALE	LEVR-FEMALE	SENS-FEMALE	(B)	LEVR/ SENS	$LEVR^{MALE}$/ $SENS^{MALE}$	$LEVR^{FEMALE}$ /$LEVR^{MALE}$	$LEVR^{FEMALE}$/ $SENS^{FEMALE}$	$SENS^{FEMALE}$ /$SENS^{MALE}$	FEMALE/ MALE
GENE ID	42880874	40243012	43172126	42179378							
Receptor Subunits											
acr-8	3.05(234)	2.36(170)	1.34(104)	1.61(122)		1.11	1.29	**0.44**	0.83	0.68	0.55
acr-16	1.21(82)	1.14(72)	0.35(24)	0.51(34)		0.95	1.07	**0.29**	0.69	**0.45**	**0.37**
acr-20	0.37(28)	0.25(18)	1.30(100)	0.50(38)		**2.20**	1.46	**3.55**	**2.57**	**2.01**	**2.92**
acr-21	0.03(2)	0.04(2)	1.38(84)	0.47(28)		**2.79**	0.94	**41.72**	**2.93**	**13.36**	**27.09**
acr-25	0.28(10)	0.29(10)	0.60(22)	0.34(12)		1.39	0.94	**2.19**	1.79	1.14	1.65
lev-1	1.06(70)	2.11(130)	1.63(108)	1.95(126)		0.66	0.51	1.53	0.84	0.92	1.13
lev-8	0.00(0)	0.00(0)	0.52(6)	0.18(2)		**2.93**	0.00	0.00	**2.93**	0.00	0.00
unc-29_{alt}	0.35(14)	0.26(10)	0.00(0)	0.00(0)		1.31	1.31	0.00	0.00	0.00	0.00
unc-29	0.06(4)	0.03(2)	0.00(0)	0.00(0)		1.88	1.88	0.00	0.00	0.00	0.00
unc-38	4.41(330)	10.77(756)	1.73(130)	4.62(340)		**0.40**	**0.41**	**0.39**	**0.37**	**0.43**	**0.42**
unc-63	0.40(38)	1.30(116)	0.25(24)	1.78(166)		**0.21**	**0.31**	0.63	**0.14**	1.37	1.19
Non-Subunit Pathway Proteins											
lev-9	0.45(80)	0.49(82)	0.65(118)	1.27(224)		0.63	0.92	1.47	0.51	**2.61**	**2.06**
lev-10	0.34(40)	1.23(136)	0.22(26)	1.11(128)		**0.24**	**0.28**	0.65	**0.20**	0.90	0.84
lev-11	434.21(20258	893.38(39116)	76.56(3596)	297.88(13670)		**0.43**	**0.49**	**0.18**	**0.26**	**0.33**	**0.28**
nra-1	7.74(806)	18.85(1842)	0.59(62)	8.20(840)		**0.31**	**0.41**	**0.08**	**0.07**	**0.44**	**0.33**
nra-2	7.61(560)	13.78(952)	43.22(3204)	33.97(2460)		1.06	0.55	**5.68**	1.27	**2.47**	**3.61**
nra-4	4.84(744)	8.05(1162)	6.82(1056)	9.72(1470)		0.66	0.60	1.41	0.70	1.21	1.28
oig-4	1.77(40)	1.61(34)	1.06(24)	1.17(26)		1.02	1.10	0.60	0.90	0.73	0.66
ric-3	2.86(162)	7.59(404)	2.49(142)	9.71(542)		**0.31**	**0.38**	0.87	**0.26**	1.28	1.17
soc-1	1.09(78)	1.43(96)	6.20(448)	3.54(250)		1.47	0.76	**5.70**	1.75	**2.48**	**3.88**
tax-6	6.04(562)	16.10(1406)	10.27(962)	18.73(1714)		**0.47**	**0.38**	1.70	0.55	1.16	1.31
tpa-1	3.00(362)	8.13(920)	8.07(980)	8.61(1022)		0.66	**0.37**	**2.69**	0.94	1.06	1.50
unc-22	4.11(3684)	8.59(7220)	0.74(664)	2.62(2306)		**0.43**	**0.48**	**0.18**	**0.28**	**0.30**	**0.26**
unc-50	16.91(1008)	15.30(856)	68.06(4084)	37.25(2184)		1.62	1.11	**4.02**	1.83	**2.43**	**3.27**
unc-68_{ty}	1.82(1218)	5.64(3548)	1.13(762)	3.14(2072)		**0.34**	**0.32**	0.62	**0.36**	0.56	0.57
unc-74	2.12(142)	2.55(160)	12.57(846)	5.75(378)		1.77	0.83	**5.92**	2.19	**2.25**	**3.92**
Reference Genes											
Ubc2	479.38(9127)	1,255.55(22434)	1,115.95(21391)	1,257.87(23557)		0.63	**0.38**	**2.33**	0.89	1.00	1.37
TPI	354.78(7911)	346.74(7256)	284.55(6388)	293.39(6435)		1.00	1.02	0.80	0.97	0.85	0.82

Expression profiles

The coding sequences for *O. dentatum* nAChR pathway genes (Table 4.1) were used as a single reference genome to map the reads from each library and thereby obtain the corresponding RNA expression profiles. Bowtie (with the e value set to the default value of 70) was used to map paired reads to the set. The number of reads mapping to each sequence was normalized to library size and sequence length to determine the expression level of a given gene in fragments per kilobase exon per million reads (FPKM). The expression levels for each sequence were compared across libraries (Table 4.2).

Nine sequences exhibited an expression difference of greater than two-fold between LEVR and SENS samples. The mRNAs for *lev-10, lev-11, nra-1, ric-3, tax-6, unc-22, unc-38, unc-63* and *unc-68$_{ry}$* were under-expressed in both LEVR male and female libraries as compared to the SENS libraries. Expression of several sequences exhibited a gender association, with *acr-16, lev-11, nra-1, unc-22, unc-29.1* and *unc-38* all exhibiting reduced expression in females, and *acr-21, nra-2, soc-1, unc-50* and *unc-74* showing increased expression in females. Of note was a disparity between the expression profiles of the two reference gene sequences; *tpi* was consistent between libraries, but *ubc-2* was under-expressed in the LEVR male library relative to the other three libraries. RT-qPCR assessment of *acr-8, ric-3, unc-38, ubc-2,* and *tpi* on the same RNA population yielded similar results (data not shown).

SNP/indel

Reads were mapped by library as paired-ends to the set of identified genes for SNP/indel assessment. The program GATK identified one coding sequence indel from among the set of 26 genes, a 6 base deletion in *nra-1* at coding sequence position 772 that was represented with nearly equal frequency in all but the LEVR female library (in which a low expression level precluded analysis).

Use of GATK for SNP calling within the combined mapped read libraries provided an initial count of 4440 possible SNPs. The 4440 count was reduced to 2856 SNPs by filtering using published criteria [12] of (1) minimum 10 reads per position, (2) at least 20% variance from the reference, and (3) 3 or more calls for an alternate allele, observed in one or both isolates. The count was further reduced to 2788 by filtering out SNPs with a GATK-called SNP quality score of <50.

The post-filter 2788 SNPs were analyzed for effect on protein sequence and difference in allelic frequency between LEVR and SENS samples. Of the 2788 SNPs, 2600 resulted in synonymous mutations, 72 were non-synonymous mutations, and 106 were in UTR's. The 72 non-synonymous SNPs and there frequencies in each isolate are shown in Table 4.3. In all, 68 of the 2788

SNPs showed an allelic frequency difference between LEVR and SENS isolates ≥ 50%, and of these, the only non-synonymous SNP was located within ACR-21 (*V272I*).

Table 4.3 – Non-synonymous SNPs of the nAChR pathway

SNPs which produce non-synonymous mutations are listed by gene and amino acid residue. SNPs with multiple alleles are starred (*). Alleles not detected in a isolate are marked nd. The frequency of the allele producing the change is also given by isolate as a proportion of the reads mapping to the site.

		Frequency of Alternate Allele	
Gene	AA Change	LEVR	SENS
Receptor Subunits			
acr-20	V141I	0.353	0.111
acr-21	V272I	0.200	0.700
	P377S	0.474	0.091
acr-8	S45A	1.000	0.952
lev-1	L45F	0.313	0.154
	T82S	0.208	0.088
	R355H*	0.762	0.944
	R355P*	0.190	0.056
unc-38	E2G	nd	0.333
Non-Subunit Pathway Proteins			
lev-9	**I400M***	nd	0.167
	Synonymous*	nd	0.167
	S591N	0.714	0.714
	T724I	nd	0.318
	A856V	0.625	0.694
	V870A	0.222	0.217
	G891A	0.208	0.200
	N913S	0.158	0.429
	H917Y	0.111	0.205
	A978T	0.545	0.471
nra-1	S9F	0.183	0.232
	I484V	0.250	0.283
	I552V	0.214	0.196
	A604V	0.271	0.260
	V609M	0.262	0.308
	P643S	0.202	0.220
	P645S	0.198	0.248
	P647A	0.205	0.047
	G649S	0.188	0.245
	A685T	0.227	0.043

		Frequency of Alternate Allele	
Gene	AA Change	LEVR	SENS
nra-2	**L21F***	0.317	0.097
	Synonymous*	0.157	0.143
	T71S	0.890	0.824
	T186A	0.137	0.224
	H219N	0.396	0.138
nra-4	N225D	0.496	0.417
	N227K	0.504	0.446
	S234N	0.545	0.522
	A239V	0.221	0.099
	A293T	0.360	0.176
	N650S	0.617	0.528
	S762N*	0.150	0.004
	S762T*	0.098	0.325
	E957D*	0.012	0.104
	E957D*	0.814	0.896
oig-4	Q33K	0.261	0.080
ric-3	E298D	0.082	0.367
	R308K	0.143	0.342
	K310R	0.292	0.198
	T313A	0.500	0.473
	A329V	0.511	0.464
	A332T	0.476	0.465
	A336V	0.303	0.186
soc-1	N103T	nd	0.231
	I311T	nd	0.265
	V314I	nd	0.229
	V345A	0.266	0.469
tpa-1	L199F	0.261	0.086
	L259F	0.318	0.014
	S401N	0.257	0.412
	L510F	0.345	0.261

		Frequency of Alternate Allele	
Gene	AA Change	LEVR	SENS
unc-22	D506E	0.273	0.500
	K2126N	0.200	0.074
	T2695I	0.500	0.450
	V2838I	0.348	0.463
	I4063V	0.125	0.275
	E4559A	0.414	0.450
	L5387F	0.235	0.272
	I6421V	0.111	0.410
	Q6428H	0.109	0.413
	T6503I	0.289	0.314
	A6745V	0.337	0.289
unc-68$_{ry}$	A811P	0.200	0.250
	V2315I	0.167	0.355
	L4131F	0.488	0.258
	E4481D	0.364	0.316
unc-74	**I14V***	0.250	0.212
	I14F*	0.182	0.303
	K41I	1.000	0.969
	Y286F	0.511	0.410

Discussion

This study reports the assessment of mRNA expression in levamisole-sensitive and –resistance *O. dentatum* strains of 26 genes of the nAChR muscle contraction signaling pathway, including 19 whose sequences were determined in silico. Hallmarks of the quality of those 19 sequences are provided by the comparison to homologs of other organism(s) (Table 4.1), i.e. with

respect to sequence length and similarity. Of note is that the *unc-68$_{ry}$* sequence determined *in silica* (Table 4.1) corresponds to the ryanodine receptor of *C. elegans* (NCBI: BAA08309). Investigators have reported the deduced protein sequences in *C. elegans* for both that ryanodine receptor and for the highly similar (>95% identity) UNC-68 receptor (e.g. NCBI: NP_001256074); their high similarity suggests likely equivalent function, calcium release from the endoplasmic reticulum. The major difference between these protein sequences is that UNC-68 (and its various isoforms) contains a 116 amino acid residue segment inserted after a residue generally corresponding to residue 2590 of UNC-68$_{RY}$. In the analysis of the sequence library contigs reported here, sequences encoding the 116 residue segment were not found. In further validation of that finding, BLAST searches that queried each of the four read libraries with the diagnostic 116-residue segment failed to identify reads that mapped. These data suggest that the homolog of the *C. elegans* ryanodine receptor, but not UNC-68, is present in *O. dentatum*.

Two-fold or greater decreases in *lev-10, lev-11, nra-1, ric-3, unc-22, unc-38, unc-63* and *unc-68$_{ry}$* expression in LEVR vs. SENS worms indicates that expression of these genes associates with resistance; reduced expression of these genes is predicted to reduce the sensitivity to levamisole. The observed reduction in expression of *unc-38* and *unc-63*, which are nAChR subunits, may explain the findings of Robertson et al. [9] showing altered levamisole receptor properties in LEVR. As RIC-3 [13], NRA-1 [14], and LEV-10 [15], are important in the proper formation of the levamisole receptor at the cell surface, decreased expression of these would likely be observed as a decreased number of functional receptors at the cell surface, which was one of the observations made in a previous study [9]. UNC-68, UNC-22 and LEV-11 are post-signaling modulators of the nAChR pathway; UNC-68 amplifies the signal [16], while UNC-22 and LEV-11 are required for muscle contraction [17]. Decreased expression of these proteins would be expected to decrease the effective response of the parasite to levamisole. TAX-6 is a phosphatase that negatively regulates the levamisole receptor [14]. An observed decrease in *tax-6* expression is more likely to be a compensatory change than causative

of levamisole resistance. The increased expression of *acr-21, nra-2, soc-1, unc-50* and *unc-74* in females may be a true gender difference in expression or may be due to expression patterns of the egg mass in females; the presence of the eggs in females suggests caution is necessary when comparing adult males and females. One observation arising from our data is that the expression profiles of the two reference genes are not equivalent; *tpi* is consistently expressed in all samples while *ubc-2* is consistent in three samples, but very low in LEVR male.

It has been noted that nematodes have a high rate of polymorphism [18, 19]. Our SNP analysis showing nearly 37 SNPs per kb supports that observation of high polymorphism. The non-synonymous SNPs, particularly those with larger frequency differences between isolates, provide an interesting starting place for protein studies, although these should be confirmed with additional data. Large frequency differences between alleles of synonymous SNPs likely are artifacts of the selection process undergone to generate the LEVR isolate. Truncations in several pathway genes have been shown to associate with drug resistance [20, 21], but no SNPs or indels causing premature stop codons leading to truncated proteins were observed in this study.

Conclusions

Based on the lack of SNPs or indels that would affect protein function, and on the observed expression levels of nAChR pathway genes, we propose that levamisole resistance in *O. dentatum* is associated with both the reduced level of receptors containing UNC-38 and/or UNC-63 subunits and with decreased signal transduction facilitated by reduced levels of LEV-10, LEV-11, NRA-1, UNC-22, and UNC-68_{ry}.

Methods

Parasite material

Adult levamisole sensitive (SENS) *O. dentatum* isolates [22] were used to infect pigs and then were harvested as described [9]. Worm counts were estimated as a function of the amount of RNA extracted. Sixty-eight male and 69 female SENS worms were collected. To remove foreign

contaminants, samples were washed at least 3 times in a pH 7.5 maintenance solution consisting of (mM): NaCl (150), KCl (2.7), $CaCl_2$ (2), $MgCl_2$ (0.3), PIPES (10), NaOH (13), glucose (11), $NaHCO_3$ (12), penicillin 0.06 g/L, streptomycin 0.1 g/L. Worms were settled by gravity before removing each wash, then blotted dry and transferred to a 1.5ml microfuge tube, weighed by difference, and processed for extraction of RNA.

RNA extraction

Parasite samples resuspended in 1.0 ml TRI reagent (Molecular Research Center) were ground by mortar and pestle under liquid nitrogen, then brought to a total volume of 2-3 ml TRI reagent. Total RNA was extracted from the TRI reagent according to the manufacturer's instructions, including an additional centrifugation step for clearing insoluble material. Extracted RNA was treated with DNase I (New England Biolabs) (10 min at 37C, 10 min at 75C), then re-extracted with TRI reagent and resuspended in diethylpyrocarbonate-treated water. RNA concentration, purity, and quality (RNA Integrity Number) were assessed on a 2100 Bioanalyzer (Agilent Technologies).

mRNA-Seq

The building of indexed, non-normalized, paired-end mRNA-Seq libraries, and subsequent 75-cycle pyrosequencing on an Illumina GAIIx platform, were performed as a service by the DNA Facility (Office of Biotechnology, Iowa State University) using 5 ug total RNA (per sample). Duplexed SENS samples were run in a single lane.

Equivalent levamisole resistant data were similarly produced [10] and are available at the NCBI sequence read archive: LEVR male (SRR393668), LEVR female (SRR393669).

Bioinformatic Components

Assembly. Velvet version 1.1.06 [23] was used for contig assembly.

Similarity searching. BLASTx, tBLASTx, and tBLASTn algorithms [24] were used to compare contigs with sequences available in public databases including the National Center for Biotechnology

Information (NCBI) to identify homologues from other nematodes, i.e. sequences returning BLAST expect values ≤ 1E-10.

Read mapping. 64 bit Bowtie [25] version 0.12.7 was used to map reads for contig building and quantitation.

BWA [26] version 0.6.2-r126 was used to map reads for SNP/indel analysis.

Pairwise Comparison. The Needle algorithm [27] was used for pairwise comparison.

SNP/indel analysis. The Genome Analysis Toolkit v1.6-7 [28, 29] was used to call SNPs and indels.

SNPeff [30] was used to determine the effects of SNP variants.

Real-time qPCR

RT-qPCR using qScript One-Step SYBR Green qRT-PCR Kit (Quanta Biosciences) was performed on a GeneAmp 5700 Sequence Detection System (Applied Biosystems) using the thermo-protocol: 95°C 5 min, and 50 cycles of 95°C 10 s, 60°C 30 s, followed by melt curve analysis. Gene specific primers were either previously published [31] or were as follows: acr-8: F 5'-CACCAATCATAAATCGCGAGG, R 5'- TTTTCCGCCGGAGAGCTAT; ric-3: F 5'-GGACGAGGAATGTTCACTTGG , R 5'- CTCCTGGAGACCACGAAGCT; unc-38:F 5'-AACTCCTCCGGCGCAGAC, R 5'-ACCTCTCCGCTTACGCGAT. Custom Excel files were used to transform quantification cycle values generated by the GeneAmp 5700 software into relative gene expression levels for each target using the efficiency-corrected standard curve-based relative quantification approach for RT-qPCR [32, 33].

List of abbreviations

FPKM: fragments per kilobase per million reads

nAChR: nicotinic acetylcholine receptor

RT-qPCR: reverse transcription quantitative polymerase chain reaction

Competing interests

The authors declare that they have no competing interests

Authors' contributions

NR participated in the design and coordination of the study, carried out the molecular genetics studies and sequence analysis, and drafted the manuscript. JG participated in the design and analysis of the RT-qPCR. RM participated in the design and coordination of the study, and helped to draft the manuscript. JB conceived of the study, participated in the design and coordination, and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Hui-Hsien Chou (Iowa State University) for assistance and Claude Charvet for the use of non-public sequence information (JX429921 and JX429919). Research funding was by The Hatch Act, State of Iowa, which had no role in the design, execution or publication of this study.

References

1. Brooker, S., *Estimating the global distribution and disease burden of intestinal nematode infections: adding up the numbers--a review.* Int. J. Parasitol., 2010. **40**(10): p. 1137-44.
2. WHO, *The selection and use of essential medicines.* World Health Organ. Tech. Rep. Ser., 2011(965): p. i-xiv, 1-249.
3. Prichard, R.K., *Anthelmintic resistance.* Vet. Parasitol., 1994. **54**(1-3): p. 259-68.
4. Geerts, S. and B. Gryseels, *Drug resistance in human helminths: current situation and lessons from livestock.* Clin. Microbiol. Rev., 2000. **13**(2): p. 207-22.
5. Prichard, R.K., et al., *A research agenda for helminth diseases of humans: intervention for control and elimination.* PLoS Negl. Trop. Dis., 2012. **6**(4): p. e1549.
6. James, C.E., A.L. Hudson, and M.W. Davey, *Drug resistance mechanisms in helminths: is it survival of the fittest?* Trends Parasitol., 2009. **25**(7): p. 328-35.
7. Martin, R.J. and A.P. Robertson, *Mode of action of levamisole and pyrantel, anthelmintic resistance, E153 and Q57.* Parasitology, 2007. **134**(Pt 8): p. 1093-104.
8. Neveu, C., et al., *Genetic diversity of levamisole receptor subunits in parasitic nematode species and abbreviated transcripts associated with resistance.* Pharmacogenet Genomics, 2010. **20**(7): p. 414-25.
9. Robertson, A.P., H.E. Bjorn, and R.J. Martin, *Resistance to levamisole resolved at the single-channel level.* FASEB J, 1999. **13**(6): p. 749-60.
10. Romine, N.M., R.J. Martin, and J. Beetham, *Computational cloning of anthelminthic target genes in the non-model nematode parasite Oesophagostomum dentatum.* 2012.
11. Blaxter, M.L., et al., *A molecular evolutionary framework for the phylum Nematoda.* Nature, 1998. **392**(6671): p. 71-5.
12. Koepke, T., et al., *Rapid gene-based SNP and haplotype marker development in non-model eukaryotes using 3'UTR sequencing.* BMC Genomics, 2012. **13**: p. 18.
13. Halevi, S., et al., *The C. elegans ric-3 gene is required for maturation of nicotinic acetylcholine receptors.* Embo J, 2002. **21**(5): p. 1012-20.

14. Gottschalk, A., et al., *Identification and characterization of novel nicotinic receptor-associated proteins in Caenorhabditis elegans.* Embo J, 2005. **24**(14): p. 2566-78.
15. Gally, C., et al., *A transmembrane protein required for acetylcholine receptor clustering in Caenorhabditis elegans.* Nature, 2004. **431**(7008): p. 578-82.
16. Maryon, E.B., R. Coronado, and P. Anderson, *unc-68 encodes a ryanodine receptor involved in regulating C. elegans body-wall muscle contraction.* J Cell. Biol., 1996. **134**(4): p. 885-93.
17. Fleming, J.T., et al., *Caenorhabditis elegans levamisole resistance genes lev-1, unc-29, and unc-38 encode functional nicotinic acetylcholine receptor subunits.* J. Neurosci., 1997. **17**(15): p. 5843-57.
18. Mushegian, A.R., et al., *Large-scale taxonomic profiling of eukaryotic model organisms: a comparison of orthologous proteins encoded by the human, fly, nematode, and yeast genomes.* Genome Res, 1998. **8**(6): p. 590-8.
19. Cutter, A.D., S.E. Baird, and D. Charlesworth, *High nucleotide polymorphism and rapid decay of linkage disequilibrium in wild populations of Caenorhabditis remanei.* Genetics, 2006. **174**(2): p. 901-13.
20. Rufener, L., et al., *Haemonchus contortus acetylcholine receptors of the DEG-3 subfamily and their role in sensitivity to monepantel.* PLoS Pathog, 2009. **5**(4): p. e1000380.
21. Williamson, S.M., et al., *Candidate anthelmintic resistance-associated gene expression and sequence polymorphisms in a triple-resistant field isolate of Haemonchus contortus.* Mol. Biochem. Parasit., 2011. **180**(2): p. 99-105.
22. Varady, M., et al., *In vitro characterization of lines of Oesophagostomum dentatum selected or not selected for resistance to pyrantel, levamisole and ivermectin.* Int. J. of Parasitol., 1997. **27**(1): p. 77-81.
23. Zerbino, D.R. and E. Birney, *Velvet: algorithms for de novo short read assembly using de Bruijn graphs.* Genome Res, 2008. **18**(5): p. 821-9.
24. Altschul, S.F., et al., *Basic local alignment search tool.* J Mol Biol, 1990. **215**(3): p. 403-10.
25. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biol, 2009. **10**(3): p. R25.
26. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.
27. Needleman, S.B. and C.D. Wunsch, *A general method applicable to the search for similarities in the amino acid sequence of two proteins.* J Mol Biol, 1970. **48**(3): p. 443-53.
28. McKenna, A., et al., *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.* Genome Res, 2010. **20**(9): p. 1297-303.
29. DePristo, M.A., et al., *A framework for variation discovery and genotyping using next-generation DNA sequencing data.* Nat Genet, 2011. **43**(5): p. 491-8.
30. Cingolani, P., et al., *A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w (1118) ; iso-2; iso-3.* Fly 2012. **6**(2): p. 80-92.
31. Ondrovics, M., et al., *Transcription profiles for two key gender-specific gene families in Oesophagostomum dentatum during development in vivo and in vitro.* Infect Genet Evol, 2012. **12**(1): p. 137-41.
32. Livak, K.J. and T.D. Schmittgen, *Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.* Methods, 2001. **25**(4): p. 402-8.
33. Pfaffl, M.W., *A new mathematical model for relative quantification in real-time RT-PCR.* Nucleic Acids Res, 2001. **29**(9): p. e45.

CHAPTER 5. GENERAL DISCUSSION

Non-parasitic model organisms typically have many experimental benefits such as ease of culture and preexistence of molecular-data (e,g, sequenced genome) and -tools allowing genetic manipulation; however, they also have many pitfalls, such as applicability to parasitic organisms. This has been demonstrated several times in nematodes with Towers et al. [1] demonstrating that *lev-8* null mutants confer partial resistance to levamisole while Boulin et al. [2] failed to show evidence of a role for *lev-8* in resistance in *Haemonchus contortus*. Additionally, a SNP in some chloride channels has been shown to cause resistance to macrolytic lactones in *H. contortus* [3] although that same SNP is not sufficient for macrolytic lactone resistance in *Caenorhabditis elegans* [4]. Thus, while research in non-parasite models may be easier it is important to translate this work into parasitic organisms.

We have described an application of mRNA-Seq for identification of target genes that greatly simplifies translation of research from model organisms to organisms lacking a sequenced genome. The first requirement for studying target genes in an organism lacking sequence information is the assembly of those gene sequences. Traditional assembly methods of second generation sequence data typically produce incomplete sequences; the method we describe improves on traditional assembly techniques by producing more complete sequences. Furthermore, this assembly method is more efficient than exhaustive k-mer assemblies as only a single seed contig of a target gene is needed for identification, drastically reducing the number of computationally intensive whole transcriptome assemblies.

We used this method to identify in the parasitic nematode *Oesophagostomum dentatum* 26 genes involved in the nAChR pathway and an additional eight genes associated with resistance to other classes of anthelminthic drugs. Given the findings of Robertson et al. [5] showing decreased response of the levamisole receptor to increasing levamisole concentration in levamisole resistant *Oesophagostomum dentatum* we would expect resistance in this isolate of *O. dentatum* to be caused

by a change in receptor sequence, stoichiometry, abundance, or post-translational modification. Our data present no evidence of SNPs that have been reported in other organisms as causative for anthelminthic resistance. Of the SNPs we report in nAChR subunits only nine were identified that would cause non-synonymous mutations. Only five of these non-synonymous SNPs were in subunits reported to be connected to levamisole resistance (*lev-1, unc-38*) with increases in allele frequency of 15.9%, 12%, 4.8%, 13.4%, and 33%. This suggests that resistance associated with altered receptor properties is caused by changes in receptor subunit stoichiometry, abundance, or post-translational modification.

We identified two-fold or greater decreases in RNA expression for several genes involved in the nAChR pathway. These include two nAChR subunits (*unc-38, unc-63*), three proteins involved in proper receptor formation on the cell surface (*lev-10, nra-1, ric-3),* a protein that amplifies signal transduction (*unc-68*$_{ry}$), and two proteins involved in muscle contraction (*lev-11, unc-2).* This implicates receptor stoichiometry (loss of receptors containing UNC-38 and/or UNC-63) and decreased signal transduction as the cause of levamisole resistance. Of note is the decreased expression of tax-6. As a negative regulator of the levamisole receptor [6] decreased expression may be a compensatory change allowing received signals to still be transduced.

A key application of this research is improving the treatment of parasitic nematode infections. In our work with a population of parasitic organisms we have identified several putative molecular mechanisms of levamisole resistance. Distinguishing molecular mechanisms of resistance to different drugs will improve drug targeting. Knowledge of the different resistance mechanisms to individual drugs can suggest drugs that are less likely to have cross resistance. Also, as the nicotinic agonist and the aminoacetonitrile derivative classes act on different families of nAChR subunits, knowledge of the mechanisms involved may show that resistance in one increases selection for sensitivity to the other. The identification of molecular mechanisms of resistance that correspond to changes in DNA will also facilitate the determination of drug resistance in the field by establishing a molecular testing

technique. Current methods of resistance testing are cumbersome and require treatment with drugs followed by a two week waiting period before assessing the level of egg shedding from the host. Increased ease of determining drug resistance will in turn improve drug selection for disease treatment. Improved treatment of infection would benefit both human health care and agricultural production.

Complications in applying the information gained here include our experimentation on a population of worms as opposed to individual worms. While it is believed that resistance is polygenic and we have shown here that many genes are associated with resistance within a population we are unable to determine from this data whether the decrease in expressions of each gene is uniform in the population. Further experimentation with individual worms would be necessary to determine whether all or a subset of the genes identified are required for resistance in a field isolate.

Additionally, the non-synonymous SNPs identified here are just those for which our sequence data provide evidence. As these sequence changes may be observed due to experimental error additional data sets will provide greater confidence in the SNPs identified and increase the power of further SNP detection. The sequence data produced for these experiments has been uploaded to the Sequence Read Archive, a publicly accessible data storage facility provided by the National Center for Biotechnology Information. The data is accessible by other groups studying levamisole resistance and may be added to by those groups to increase the data depth.

Future Directions

This work was performed with the intended goal of identifying full length protein coding sequences for genes of interest involved with levamisole resistance. In future work, these sequences will be expressed in a *Xenopus laevis* oocyte system for functional analysis. Identification of full length coding sequences for most of the genes explored will facilitate the reconstitution of functional nAChRs. It is necessary to assess the electrophysiological properties of these reconstituted receptors in response to anthelminthics to understand variations between nAChR subtypes. These different

subtypes respond differently to different drugs while allowing the organism to maintain normal function. The same *Xenopus* reconstitution methodology will allow the assessment of SNPs on receptor function/sensitivity (to levamisole), although it is important to first experimentally validate these non-synonymous SNPs.

Another ongoing study that issues from this thesis work is that, in collaboration with the ISU Genome Informatics facility, we have begun initial steps toward application of this approach on a larger scale to facilitate assembly of a complete de novo *O. dentatum* transcriptome. This will provide a global view of the RNA population. As anthelminthic resistance is polygenic, a global view will help identify other genes associated with levamisole resistance. A complete transcriptome is also necessary for identification of compensatory changes that serve to allow resistant worms to function normally even though their nAChR activity is reduced.

Many other avenues of research are opened by the research detailed here. Identification of these genes and the process by which they are identified allows for comparative genetics studies among nematode species that will help to understand differences in drug sensitivity between species. The RNA expression patterns seen here also provide a target profile that may be used to identify levamisole resistance in field isolates. Monitoring of field populations for expression changes seen here can be used to determine when a population has developed resistance to levamisole to determine when it is necessary to switch anthelminthic classes.

With widespread resistance the emphasis on drug treatment of infected animals has shifted from frequent anthelminthic treatment of all animals to treatment of only the most heavily infected animals. In many organisms drug resistance has a fitness cost. This suggests that with the relief of drug pressure negative selection on resistance genes would result in a population shift to drug sensitivity. If a time point is established at which resistance is lost it would provide incentive to halt treatment with levamisole to allow field populations to lose resistance. This, in turn, would improve the effects of targeted treatment of individuals with high parasite burdens. Using the RNA expression

profile of resistance associated genes established in this work it would be possible to relieve the drug pressure on a resistant isolate and monitoring the population through many passages. Establishing a time point at which resistance is lost would be extremely useful information for treatment of parasitic nematodes.

References

1. Towers PR, Edwards B, Richmond JE, Sattelle DB: **The Caenorhabditis elegans lev-8 gene encodes a novel type of nicotinic acetylcholine receptor alpha subunit**. *J Neurochem* 2005, **93**(1):1-9.
2. Boulin T, Fauvin A, Charvet CL, Cortet J, Cabaret J, Bessereau JL, Neveu C: **Functional reconstitution of Haemonchus contortus acetylcholine receptors in Xenopus oocytes provides mechanistic insights into levamisole resistance**. *Br J Pharmacol* 2011, **164**(5):1421-1432.
3. Njue AI, Hayashi J, Kinne L, Feng XP, Prichard RK: **Mutations in the extracellular domains of glutamate-gated chloride channel alpha3 and beta subunits from ivermectin-resistant Cooperia oncophora affect agonist sensitivity**. *J Neurochem* 2004, **89**(5):1137-1147.
4. Dent JA, Smith MM, Vassilatis DK, Avery L: **The genetics of ivermectin resistance in Caenorhabditis elegans**. *Proc Natl Acad Sci U S A* 2000, **97**(6):2674-2679.
5. Robertson AP, Bjorn HE, Martin RJ: **Resistance to levamisole resolved at the single-channel level**. *FASEB J* 1999, **13**(6):749-760.
6. Gottschalk A, Almedom RB, Schedletzky T, Anderson SD, Yates JR, 3rd, Schafer WR: **Identification and characterization of novel nicotinic receptor-associated proteins in Caenorhabditis elegans**. *Embo J* 2005, **24**(14):2566-2578.

APPENDIX A. SCRIPT FOR TARGET GATHERING FROM BLAST OUTPUT

```
import java.io.*;
import java.util.*;

public class BLAST_Outfmt6_Subject_ Lookup
{
	/** Gathers contigs from source files that represent the genes of interest based on BLAST
		analysis
	 * Length of match and identity are adjustable in getSeqs method, e value is set by BLAST
	 *
	 * @param args[0] the file name containing the BLAST output in format 6
	 * @param args[1] the file containing the source contigs
	 * @param args[2] the desired output file name
	 */
	public static void main(String[] args) throws IOException, FileNotFoundException
	{
		//Get file names from command line
		String seqFileName = args[0];
		String srcFileName = args[1];
		String outFileName = args[2];

		File seqFile = new File(seqFileName);   //file containing the BLAST output
		File srcFile = new File(srcFileName);   //file containing original contigs

		String[][] output = getSeqNames(seqFile);          //get list of contigs from BLAST
			that meet criteria
		System.out.print("Got the sequence names!\r\n");
		output = getSeqs(output,srcFile);          //search original contigs file to associate
			contigs with names
		System.out.print("Got the original sequences!\r\n");
		System.out.print("Printing.  Almost done... ");

		File myFileOut = new File(outFileName);                  //output file
		if(!(myFileOut.exists()))                                //Make sure file exists
		{myFileOut.createNewFile();}
		FileWriter out = new FileWriter(myFileOut);     //Open an output stream for the file
			specified

		for(int g = 0;g < output.length;g++)       //print output array in FASTA format
		{out.write(">"+output[g][0]+"\r\n"+output[g][1]+"\r\n");}
		out.close();                                             //close output file
		System.out.print("Done!\r\n");
	 }

	/* Collects contig sequences that match names gathered in previous method
	 *
	 * returns array of names and sequences
	 *
	 * @param outP 2d string array for contig names and sequences
	 * @param srcFileP contig source file
```

```
 */
private static String[][] getSeqs(String[][] outP, File srcFileP)throws
        FileNotFoundException
{
        Scanner srcFileIn = new Scanner(srcFileP); //file reader for original contig file
        int numFound = 0, last = 0, numAdded =0;
        String check = "", seq = "";
        while(srcFileIn.hasNext() && numFound <= outP.length)
        {
                check = srcFileIn.nextLine();           //grab next line
                if(check.contains(">"))                         //if line is a sequence name
                {
                        if(numAdded < numFound) //place current sequence to next location
                        {
                                outP[1][1] = seq;
                                numAdded++;
                        }
                        seq = "";       //clear sequence
                        for(int i = 0;i < outP.length;i++)//scan output array to see if matches
                        contig target
                        {
                                if(outP[i][0].equals(check.substring(1))) //if names match,
                                        grab sequence
                                {
                                        numFound++;   //found a new sequence
                                        System.out.print("I found "+numFound+" original
                                                sequences.\n");
                                        last = i; //mark this location in the output array
                                }
                        }
                }
                else //if not a sequence name continue building sequence length(source files
                        are 60 char limit)
                {seq = seq + check.replace("\n", "");}     //remove newline character
        }
        srcFileIn.close();
        return outP;
}

/* Pull names of sequences that have BLAST matches from input file
 *
 * @param seqFileP BLAST output file, format 6
 * return array of sequence names
 */
private static String[][] getSeqNames(File seqFileP) throws FileNotFoundException
{
        Scanner seqFileIn = new Scanner(seqFileP);  //file reader for BLAST file
        Boolean found = false;  //check to see if already found this sequence
        String[] check;         //temporary variables for holding lines from file
```

```
        String[][] out = new String[10][2];         //array for holding output data
        int seqLength, lengthEnd, regionLength;
        int curr = 0;
        while(seqFileIn.hasNext())                              //reading file
        {
                check = seqFileIn.nextLine().split("\t");  //split line containing BLAST HSP
                if(Float.parseFloat(check[2]) < 60.0) //*****IDENTITY
                        THRESHOLD*****
                {continue;}
                seqLength = check[1].indexOf("length_"); //*identify contig length
                lengthEnd = check[1].indexOf("_", seqLength+7);
                seqLength = Integer.parseInt(check[1].substring(seqLength+7, lengthEnd));
                        // end identify contig length*/
                regionLength = Integer.parseInt(check[3])* 3; //identify length of HSP match
                if(seqLength/2 > regionLength) //*****LENGTH OF MATCH
                        THRESHOLD*****
                {continue;}
                found = false;                          //only grab each sequence name once
                for(int i = 0;i < curr;i++) //scan to make sure not already found this sequence
                {
                        if(check[1].equals(out[i][0]))
                        {
                                found = true;
                                break;
                        }
                }
                if(found) //skip contigs if already found
                {continue;}
                if(curr >= out.length) //if array full
                {out = resizeArray(out,2*out.length,curr);}         //double array size
                out[curr][0] = check[1]; //write successfully blasted seq name to array
                curr++; //increment counter
        }
        out = resizeArray(out,curr,curr);           //slim output array to exact size
        seqFileIn.close();
        System.out.print("There were "+curr+" unique contigs.\n");
return out;
}

/* Resize a 2d String array
 *
 * @param oligosP the output array
 * @param max the size of the new array
 * @param actual the size of the current array
 *
 * return new sized array of sequence names
 */
private static String[][] resizeArray(String[][] oligosP,int max,int actual)
{
```

```
                String[][]tempArray = new String[max][oligosP[0].length];        //create new array
                for(int i = 0;i < actual;i++)                                    //copy the old array
                {
                        for(int j = 0;j < oligosP[0].length;j++)
                        {tempArray[i][j] = oligosP[i][j];}
                }
                return tempArray;
        }
}
```

APPENDIX B. SCRIPT FOR GATHERING MAPPED READS AND PAIRS

```
import java.io.*;
import java.util.*;

public class Bowtie_pair_lookup
{

	/** Gathers all reads that map to a given "genome" from Bowtie output file
	 * File should be produced by Bowtie using shuffled paired reads in unpaired format.
	 * This will keep reads in paired order, but map reads where only one end aligned
	 *
	 * @param seqFileName is the alignment file produced by Bowtie
	 * @param srcFileName is original paired reads file (input for Bowtie)
	 * @param outFileName the desired output file name (no file extension)
	 */
	public static void main(String[] args) throws IOException, FileNotFoundException
	{
		//Get file names from command line
		String seqFileName = args[0];			//Bowtie alignment file
		String srcFileName = args[1];		//PE shuffled reads file
		String outFileName = args[2];		//output file name, exclude extension

		//Use designated file name
		File seqFile = new File(seqFileName);	//file containing seqs mapped by Bowtie
			(paired input, mapped unpaired)
		File srcFile = new File(srcFileName);	//original reads used as Bowtie input

		String[] names = getSeqNames(seqFile); //get list of seqs mapped
		String[] seqs = new String[names.length], values = new String[names.length],
			pairNames = new String[names.length], pairSeqs = new
			String[names.length], pairValues = new String[names.length];	//build data
			arrays, will be same order as names
		System.out.print("I have the read names.\r\nThere were "+names.length+" reads
			mapped by Bowtie.\r\n");
		getSeqs(srcFile, names, seqs, values, pairNames, pairSeqs, pairValues); //search
			original reads file for seqs
		System.out.print("I got the pair sequences!\r\n");
		System.out.print("Printing.  Almost done");

		//Make sure files exist and open for editing
		//Open an output stream for the File specified
		//FASTA
		File fastaOut = new File(outFileName+".fasta");
		if(!(fastaOut.exists()))
		{fastaOut.createNewFile();}	//make sure file is available to write
		FileWriter fasta = new FileWriter(fastaOut);
		//FASTQ
		File fastqOut = new File(outFileName+".fastq");
		if(!(fastqOut.exists()))
		{fastqOut.createNewFile();}	//make sure file is available to write
```

```
                FileWriter fastq = new FileWriter(fastqOut);
                for(int i = 0;i < names.length;i++)          //print reads
                {
                        System.out.print(".");
                        fasta.write(">"+names[i].substring(1)+"\r\n"+seqs[i]+"\r\n"); //print read
                        fastq.write(names[i]+"\r\n"+seqs[i]+"\r\n"+"+"+names[i].substring(1)+"\r\n"
                +values[i]+"\r\n");         //print read
                        if(pairNames[i].equals("pair mapped"))  //if pair was mapped, next,
otherwise print pair info
                        {continue;}
                        else
                        {
                                fasta.write(">"+pairNames[i].substring(1)+"\r\n"+pairSeqs[i]+"\r\n");
                                 fastq.write(pairNames[i]+"\r\n"+pairSeqs[i]+"\r\n"+"+"+pairNames[i
                                        ].substring(1)+"\r\n"+pairValues[i]+"\r\n");
                        }
                }
                fasta.close();
                fastq.close();
                System.out.print("Done!\r\n");
        }

        /** Collect names of sequences from input file
         *
         * @param seqFileP file containing shuffled paired end reads
         * returns array of sequence names
         */
        private static String[] getSeqNames(File seqFileP) throws FileNotFoundException
        {
                Scanner seqFileIn = new Scanner(seqFileP);  //file reader
                String check = "";          //temporary variable for holding lines from file
                int set = 0, curr = 0;      //counters for position and number reads
                String[] names = new String[10];          //array for holding output data (read names)
                while(seqFileIn.hasNext())        //go through file
                {
                        set++;   //keeps track of position in entry
                        check = seqFileIn.nextLine();     //read file
                        if(set == 1)       //first line is name, 2 is seq, 3 is name, 4 is data
                        {
                                if(curr >= names.length)           //if array full
                                {names = resizeStringArray(names,2*names.length,curr);}
        //double array size
                                names[curr] = check;
        //grab name
                        }
                        else if(set == 4) //if finish entry
                        {
                                set = 0; //reset state
                                curr++; //mark another read done
```

```
                }
            }
            seqFileIn.close();
            return resizeStringArray(names,curr,curr);        //minimize array size
        }

        /**     Copy array to new array of desired size
         *
         * @param string current array
         * @param max new array length
         * @param actual current array length
         *      Returns new array of desired size containing same elements
         */
        private static String[] resizeStringArray(String[] string,int max,int actual)
        {
            String[]tempArray = new String[max];   //larger array
            for(int i = 0;i < actual;i++          //copy old array
            {tempArray[i] = string[i];}
            return tempArray;
        }

        /**
         * Collects seqs that match names gathered earlier
         * If both pairs did not map associates paired read with the one that did
         *
         * Input file must be shuffled
         *
         * @param srcFileP paired end shuffled file used as input for Bowtie run
         * @param names mapped sequence name
         * @param seqs mapped sequence
         * @param values mapped sequence quality scores
         * @param pairNames mapped sequence pair name
         * @param pairSeqs mapped sequence pair sequence
         * @param pairValues mapped sequence pair value
         * operates on filling arrays of names, sequences, and values plus pairs that matches where
need be
         */
        private static void getSeqs(File srcFileP, String[] names, String[] seqs, String[] values,
String[] pairNames, String[] pairSeqs, String[] pairValues)throws FileNotFoundException
        {
            Scanner srcFileIn = new Scanner(srcFileP);        //file reader
            System.out.print("Searching for read sequences.\r\n");
            int numFound = 0;        //Progress counter
            String check = "", prev = "", prev2 = "", prev3 = "", prev4 = "";    //temp variables
            while(srcFileIn.hasNext() && numFound < names.length)          //look for seqs, until
        find all mapped reads
            {
                prev4 = prev3;
                prev3 = prev2;
```

```
                    prev2 = prev;
                    prev = check;
                    check = srcFileIn.nextLine();     //read file, storing previous lines
                    if(check.contains("@")) //only read work from lines that start an entry
                    {
                              if(names[numFound].equals(check))      //if at line with name
compare name at the numfound-th position
                              { //since Bowtie maintains read order can search in order
                                        if(numFound > 0 && check.substring(0, check.length()-
1).equals(names[numFound-1].substring(0, names[numFound-1].length()-1)) || numFound <
names.length-1 && check.substring(0, check.length()-1).equals(names[numFound+1].substring(0,
names[numFound+1].length()-1)))                   //look to see if pair was matched
                                        {pairNames[numFound] = "pair mapped";}        //mark if
pair mapped or not
                                        else{pairNames[numFound] = "pair not mapped";}
                                        if(check.charAt(check.length()-1) == '1')//if this is the first
of the two paired reads
                                        {
                                                  seqs[numFound] = srcFileIn.nextLine(); //grab seq
                                                  srcFileIn.nextLine();      //burn quality ID
                                                  values[numFound] = srcFileIn.nextLine(); //grab
quality data
                                                  if(pairNames[numFound].equals("pair not
mapped"))        //if pair not mapped
                                                  {
                                                            pairNames[numFound] =
srcFileIn.nextLine();      //grab pair name
                                                            pairSeqs[numFound] = srcFileIn.nextLine();
//grab pair seq
                                                            srcFileIn.nextLine(); //burn pair quality ID
                                                            pairValues[numFound] =
srcFileIn.nextLine();      //grab pair value
                                                  }
                                        }
                                        else if(check.charAt(check.length()-1) == '2')      //if the
second of two paired reads
                                        {
                                                  if(pairNames[numFound].equals("pair not
mapped"))        //if pair not mapped
                                                  {
                                                            pairNames[numFound] = prev4; //grab pair
name
                                                            pairSeqs[numFound] = prev3; //grab seq
                                                            pairValues[numFound] = prev; //grab pair
name
                                                  }
                                                  seqs[numFound] = srcFileIn.nextLine();
          //grab seq
                                                  srcFileIn.nextLine(); //burn quality ID
```

```
                                                values[numFound] = srcFileIn.nextLine();
        //grab quality data
                                                if(numFound < names.length-1 &&
check.substring(0, check.length()-1).equals(names[numFound+1].substring(0,
names[numFound+1].length()-1)))         //if first pair read mapped after second.  Needed as source
file would have already passed
                                                {
                                                        seqs[numFound+1] = prev3; //grab seq
                                                        values[numFound+1] = prev; //grab value
                                                        pairNames[numFound+1] = "pair mapped";
//note that pair was mapped
                                                        numFound++;   //increment counter
                                                        System.out.print("I've found
"+numFound+"/"+names.length+" read sequences\r\n");
                                                }
                                        }
                                        numFound++;   //increment counter
                                        System.out.print("I've found
"+numFound+"/"+names.length+" read sequences\r\n");
                                }
                        }
                }
                srcFileIn.close();
        }
}
```

APPENDIX C. MRNA SEQUENCES FOR ANTHELMINTHIC TARGET GENES

>lev-1
GGATGGGCTCGATGATGACAGACGCGGAGGATCGGCTGATGGTGGACCTCTTCCGCGGC
TACAACTCGCTGGTCCAGCCGATTCGCAACAAAACCGATCTGCCGATGATCATCAAAAT
CGCCATGCAGCTAGTTCTACTTATCAATGTGGACGAGAAAGAGCAAGTAATGCACACAA
ATGTATGGCTCACTTTGAAATGGCAGGACTTCCAACTGCAGTGGGAGCCGAACGATTAC
GACGGAATAACGCAGATTCGAGTGGCTCCGGACAAGATTTGGCTGCCGGACATCGTTCT
CTTTAATAATGCTGACGGCAATTATGAGGTGTCGTTCATGTGCAACGTGCTCATACATCA
CTCTGGAGAGGTCCTCTGGGTGCCTCCAGCCATCTACAAGAGCAGCTGCATTATTGATGT
CGAATTCTTTCCTTTCGATGACCAGCGATGCAGTTTAACTTTCGGCAGTTGGACATACAA
CCGTGACGAGGTTAAGCTCGATTTTCTACATTCGGACCGCGTTGACTTCTCAGAGTATGC
TACTTCCTCAATCTGGGATATGATGGATGCGCCGGCAGTTCTCACTGAAGACAGGAGTCG
GATTGAATTCCAAGTTAGGATCAGGAGAAAAACCCTGTTCTACACTGTTGTTCTTATTCT
TCCTACTGTGCTTATGGCCTTCTTAAATGTTACTGTGTTCTACCTACCTACTGCTTCAGGC
GAGAAGATGGGTCTGACGATGAACGTGCTGCTGTCCATTGTTGTGTTCCTGCTGCTGGTG
TCCAAGATCCTCCCACCCACCTCCTCCTCGATTCCGCTGGTCGCCAAGTACCTCCTGCTCA
CCTTCGTGCTCAACATCATCACAATCCTCATCACTGTGATCATCTGCAACATCTACTTTCG
CTCCCCCATCACTCACAGACTGCCTCCGTGGGTGCGACTCGTCTTTCTAGATTGGTTGCCT
CTTCTGATGTGCATGCAGCGCCCGAGGAGGAAGAACGTTCTGCACCAGAAGAAGAAGAG
ACTGCACGACACGATGTCGAGCCATCAACTGGCGCCGCAAAATCATCACCCACACTGCA
AGGCCGCCGATAACGTCAGGGACGGCGCACCGTTGATTCGAATTAGCCAGCCGTCAGTT
GAAGAGTTGTCACCAGATGCTCAGCGAGCTGTCGACGCCATCGAATTCATCACCGAAAA
CATGAAAGATGATGAAATTACAAAGCAGTATCGAGATGACTGGAAATTCGTGGCAATGG
TCGTGGATCGAGTGTTGCTATACGGTTTCTTCGGAATAACCCTTGGCGGAACGATAGGCA
TTCTCTTCTCTGCTCCGACGGTTTTCGAGCGAGTAGACGAGCACAAGCATCTGCAGAAGC
TAATACAACTGTACAAACAAGGTCTTCCCGTGAACGACACCTTTATTATGCCGCCTTATT
AACTGTTTTTGGTGCTGAATATCTTTAAGGGCAGCGTACGCAGAAATTTTTTATGTTTTGG
AGTGCTAATGCAAAGCGGCAGAGAGTGTAGGGTTCTTCTAGGAAGTTGAACTTGCACGT
CTTATGGTGTGGAGAGTTTAAATACGAAAAAGCAAAGTGTATTATCCAAATTTCCCGG

>lev-8
CCGAATGGGACCTGCTCCGCGCAACCTCGTCGCGGTATGCACAAATTTACCCCGGATGCT
GTGGACAACAGTACTACATCGATATTCGCTATAATATTGTGATACGCAGAAAAGCGATA
TTCTTCACCGTCATGCTCACAATACCTTGCATGCTAATAGCGAACTTAACTCCATTCGTCT
TCGTTATTCCACCAAATGAGCATAAAATGACGTTTTCCATTTCTGTTTTCGTCGCGTTCAC
TCTGTTCTACCTCGTACTGATAGAACTC

>unc-29
ATTCAAGTTGTACTGGCCGTCCATACAGCGAAAGTGCTCGCCTCCGACGATGAAGAACG
CCTTATGATGGACGTGTTCCGTGGCTATAACAACCTTATTCAGCCGGTGAAAAATATCAG
CGACACGCCGATTATTGTGAAAATCGCGCTACAACTGGTGCTTCTTATAAATGTGGATGA
AAAGGACCAGGTAATGCACACCAACGTTTGGTTGACACTGAAGTGGCACGATTTCCAAA

TGCGTTGGGATCCGGTCAACTACGGAGAGATCAATCAGATTCGCGTTTCACCGGATAAA
GTCTGGCTACCCGATATTGTGCTGTTCAACAACGCTGACGGCAACTACGAGGTATCCTTC
ATGTGTAATGTCGTCATAAACAATCTTGGGGATATGCTCTGGGTACCACCGGCGATCTAC
AAAAGTTCCTGTATAATAGATGTAGAATTCTTTCCCTTCGACGAGCAAGTATGTACACTG
GTGTTTGGATCTTGGACATACAATGAAAACGAGATCAAACTCGAGTTCGAGCAAGCAGA
GTGGGTGGATCTCTCTGAATATGCGCCTTCCTCAATATGGGATGTGATCGATGCGCCTGC
TTCGCTGGTGAACAAGCG

>unc-38
CAGCTGCGGCGACCTGTGGTGCTCAAGTCTCCCTTCCCCAGTTGTTCATGTGATAGCTGG
TTGCAAGTGCGTATTATGGAGGCCTGTTTAATTCTTCTGCTTCTTTTTCTTTTGTCGAATC
ACCTTGTTTCTGGAAATGAAGACGCAAAGCGGCTGTATGACGATCTCATGGTAGACTAC
AATCGACAAAAGAGACCTTCTCTTGGGCCACACAAACCTATAACAATCAAGCTGAAGCT
AAGGTTGAGCCAAATAATAGACGTGCACGAAATCGACCAAATTATGTCGTGTTCTGTTTG
GCTGAAGCAGGTGTGGATGGACAAAAGACTGTCGTGGGATCCTAAAAATTATGGTGGTG
TCAGTGTACTTTATGTGCCTTATGAGATGATCTGGGTACCTGACATTGTTCTCTATAACAA
CGCTGATAGCTATTATAACATCACTATCTCAACGAAGGCGACTCTGCATTATTCTGGCCA
AATCACATGGGAACCCCCTGCAATTTTCAAGAGCATGTGTCAGATTGATGTTAGATGGTT
TCCTTTCGATGAGCAACAGTGCTTTATGAAGTTCGGCTCATGGACGTACTCTGAAAGCCT
CCTTAATCTGGAACTACTTGATGAGAACGTACGCTATCAAGAGGAAGTGAATGAGCAAG
GCATTGTTGACAATATAACGATAGCAGAGGACGGAATAGATTTGTCTGATTACTATCCTT
CTGTAGAATGGGACATAATGTCTAGAGTGGGCATACGCCATACAAAAAACTATCCCACG
TGCTGTAAGGACAGCGCTTACATTGATATCACGTACTATTTAAATCTTCGGCGCAAACCG
TTGTTTTACACGGTTAACCTAGTGTTTCCTTGCGTTGGTATATCATTCCTGACAATCGCTG
TGTTCTATCTGCCCTCGTATTCCGGTGAAAAAGTTTCACTTTGCATATCAATTCTTGTTGC
GTTGACTGTCTTCTTCCTTCTGTTGATTGAAATAATACCAGCAACTTCCGTTACCTTACCG
CTAATTGGGAAATATTTGGTGTTCAATATGATCATGGTTACTTTAAGCGTAATTGTCACC
ATTGTCTCTCTGGGTATTCATTTTCGAACTCCTCCGGCGCAGACAATGCCGAAATGGGTG
AAACTAGTGTTTTTAATTTGGTTGCCGAAGCTTTTAATGATGGAGCGACAACCAGATGAA
CAGGAGGTGTCTCTGAGAAGAGTGAGCCCAAGGAAGGGATTCGAGCCTAACTCTCTAGA
TGGCAAGATTCCTCTCAATTATCATCAACATCGCGTAAGCGGAGAGGTACCTATTGATGA
AAGGATACAAAAGCTTTACTATTCACCTCAGGTGATAAAAGCTTTCGAAAATATCGTTTT
GATTGCTGAAATGCTCAAGAAGAATGAACGTGACGATAAGCTGGAAGAGGATTGGAAG
TATGTTGCTGCCGTGTTGGATCGGTTCTTTCTTTTTGTCTTCTCCATTGCATGTATGTCTGG
GACTTTAGTGGTCCTCCTCCAAGCACCAACACTTTATGACTACAGGAAGCCGATCGATTT
GCAGTATCGTCCGGCCAATTTGAGTGCACTTTTTGAATGAGTATATTAGGCCAATTTTGT
TGAGAAGTATCG

>unc-63
CCGACGACTACGGTGGTGTAGACGTCCTGTATGTGCCTTCCGACACCATATGGCTTCCTG
ATATTGTGCTTTACAATAACGCTGATGGCAACTATCAGGTCACAATTATGACCAAGGCAA
AGCTCTCCTACAATGGCACGGTCGAATGGGCTCCACCTGCTATTTATAAGAGCATGTGCC
AAATAGACGTCGAGTTCTTCCCTTTCGACCGACAACAATGCGAAATGAAATTTGGTTCAT

GGACGTACGGCGGTTTGGAGGTGGACTTGATACATAAGGACGAGCATCTACAACAGGAG
ATGATCGAGATTGTCGAAGGCGTCGACGGGCCAGTGGAAGAGTCGGTCTGGATAGTTGA
CGAAGGTATTGATTTGAGCGATTACTATCCCAGCGTTGAATGGGATATTCTCAAAGTGCC
GGGTAAACGACACTCTAAACGGTATCCTTGCTGCGAGAGTCCCTTCATAGACATCACCTA
TGAAATTCATCTCCGGCGGAAAACGCTTTTCTATACTGTAAATCTTATTTTTCCTTCAGTA
GGCATAAGTTTTCTCACAGCCCTCGTCTTTTACTTACCGTCAGATGGAGGCGAAAAAATA
TCACTCTGCATCTCTATTCTTATTTCCCTCACTGTGTTCTTCTTGTTATTGGTTGAAATCAT
TCCGTCGACTTCTCTTGTCATACCTTTGATAGGCAAATACCTTCTCTTTACTATGGTGCTG
GTCACTCTCAGCGTCGTCGTCACTGTCGTTACATTGAACGTTCACTACCGATCGCCGACA
ACTCATAACATGCCTGAATGCGTTCGGCACTTCTTCATAGAAGTTCTTCCAAAGTACTTG
TTAATGAAGCGCCCTCCACAACCCGGAAGCAAGCAAAAGAAAGATTCCCGGACGACGTT
TGGTACACCACACGGGAACGGCCAATTCGCCAGCGCCTTGGCAGACCAAAATCTCGACC
TTTTATCGCCGACATTTCGCACATCTTTTGCCACGGGGTCTGCTGAAGAGGCATCGTATT
CATCATTCCAGCGAGAGCTCTCTCCTGTACGGTCAGCTGTGGAGAGTGTTGCATATATTG
CGGATCATCTCAAGAACGAAGAAGATGACAAACAGGTGATAGAAGATTGGAAGTACAT
CTCAGTGGTTATGGATCGAGTGTTTCTGCTGCTTTTCACTTTCGCTTGCGCGTTGGGCACT
ATTCTGATTATCGCCCGAGCGCCTTCAATATACGACACTACGGTACCTCTGGCATGAGCG
AGTAGCCAGCTGGGTAGCACCGCATTTCGGAGTCGTGGTAATGCCATGACACTCGTCTTC
TGCCTTTCATGTTTAATTCCACTAGGGTCGTCGTCTACGGGTTTTCATATCCTGCCAATTT
CTCACTGTTTGTCTTCGGAGTTGATTGCGATTGCGCGCAACGGAGGAGCTTTCTTTCTAAT
ATTTCTGCAATACTTTCCTACCTGCTGTGTCGGCCCCAACCTGTTTTGCCAATTATTCTTT
AAACCTGCGTTCCGGAAGGAACCATCCCCTAGTTTAACTAAAAATTTGCTGCCAATGCTG
TTGCCTTGAAGTCTACCCTCTTACGACTCCCTCGATCATAGGCTTATTCACATTTGCTCAA
CACTTCATCTTTCGTTAAATAACAATGGCGCGACACACACAGTCTGAAATAACTAGGTAA
CAATTTGTTTTGTGAAAGATATCTATTTTTTTACCTGTTATTTGTTTCTAATAAAATTCATA
GTTAGGTGTAGATTTTCATATCGTC

>avr-14
GGATAGCTGGTTGGCAGCAAATCACGAAGCAGCCTGTCGAGGTTCAGGAGCGCACTTAT
GTGACACAATAGAATGCTGCTGGCGGTACTCGCACTCGTGCTCGACGTGGCAAGGAGCA
GCGAAATCGGGTATGAAAAACTACTCGATGAGCAGAAGATCATTAAGAACTTACTGGAG
AACCCACGCAACGATTACGACTGGCGGGTTCGCCCAAGGGGGCGTCTCGATGAGTCGGA
CTTTGATTTTGTCGACCCTGAACCGGTAGTCATTACGGTCAATATGTACCTCAGAAGCAT
TTCCAAAGTTGATGACGTCAACATGGAATACTCACTGCATTTTACGTTCCGAGAAGAATG
GACGGATGAGCGGCTCTATTTTAATAGTCCCTCCTTGAAGCATATCGTGCTTTCACCAGG
GCAGAGGATCTGGGTCCCCGACACTTTCTTCCAAAACGAAAAGGACGGCAAGAAGCATG
ACATCGATACCCCGAATATCCTCATCCGAGTATATAACGGGACCGCCCGAATACTGTACT
CCTGTCGGCTAACGCTCACACTTAGCTGTCCGATGAAGTTAGCAGACTATCCTCTCGATG
TACAAACATGCGTAGTCGATTTCGCATCCTACGCCTATACTACGAAAGATATAGAATATA
TTTGGAAAGAGGAGAAACCGATCCAGATAAAAGACGGCCTCCGCCAATCACTGCCTTCG
TTTTTGCTCAGCAATGTACGAACTGGGAACTGTACATCAGTTACAAATACAGGTGCCTAT
TCCTGCCTCCGAACGATAATAGAGCTAAAAAGGGAGTTCAGTTACTATCTGTTGCAGTTG
TACATACCTTCATTTATGCTTGTAGCTGTATCTTGGGTATCATTTTGGCTAGACAAGGACT

CAGTCCCTGCTCGAGTAACTCTCGGAGTGACAACGCTCCTAACGATGACCACCCAAGCCT
CTGGTGTAAACGCCAACCTTCCGCCGGTTAGCTATACCAAAGCCATAGATATATGGATTG
GTGTCTGTTTAGCATTTATCTTCGGCGCTTTACTGGAATTCGCCTTGGTGAACTGGGCAGC
TCGCCAGGATCTCGCCATGAACACGATGAGGATACGAGCACAGCAGCAGAGCCTACATA
TGTTCTTTAGAAACGGCCGTGACGCACATGTTCGGCGCTGCCAGTATGGTGATGTCACTC
ATCTAGTCAATGGATATAATTCGATAGAGAGAAACCAGCATATGCGCCGATGTGAACCT
CACGCTTATGCTGCTATACCGAATGAGCATTTCTTCGAAGAATTGCCCTACAATTGGTGG
GACAAACTTTGGCGGATTAAATACAAGGAGCACAGTAGACGTATTGATCTCATATCTCG
AGTAATGTTCCCCTCGTTCTTCATGATTTTTAATATTACGTACTGGTGGCGCTACCTGACT
CCATACTTGGCAGTTCAAGCGCAGCTGGACTAACTCAGGCTGCTGTGGCCG

>avr-15
CAAGTTTGAGTTTTGTGGCGACCGACTAGACAGCCAGTTTGCAGCTATCAATGCAATGCA
TGGTCTCCTAATTGTGGCAGCTCTTCTTCTGAGGCTTGTGCAATGTGATCGAAATGCTGC
ACAGCTGAGCGAAAGCGAAAGACGATCTCCTAAAATGAAGCCTGCGGACATCTTCGCTA
AAATGCTTAACTCCGGCTATGACAGACGAATACGACCCCCAAATCGCGATGAACACGGA
AAGAACAAGCCTGTCGTTGTTGATGTGAACTTCTACTTAAGAAGCATATCGAACATTGAC
TTCGTGAGGATGGAGTACAGTCTGCAAATCACATTTCGACAGTTTTGGCACGATCGTCGC
CTCGAATATGGGTCCATGTTTAAAGGAAGAGAAGTGCCGAAGTTCCTGATACTGACTGA
CAAGGACGCCATCTGGACCCCGGACACATTTTTCATGAATGAGAAGCGCGCTCATCGCC
ATGATATCGACAAGCTGAACCTTATGATACGGGTTCATCCGAATGGAACCGTAATGTACT
CAGAGCGGATATCTCTTGTCCTTTCTTGTCCAATGTATATACAGAATTATCCTATGGATG
AACAAGTTTGTGGCTTGGACCTTGGTTCCTACGCCTTTACGACGGATGACATACTATATC
ACTGGCACGATCCAAATCCGATTCAGTTTCATCCTCTGCTCAATACCTCACTGCCTAGTTT
TATCATTCGCCAGGCTTTTACTGACACCTGCTCTAGCCTAACATCTACGGGAGAATATTC
CTGCATTCGGATGGTTCTTCATCTGAAAAGGCTTTTCAGTTACTACATGGTTCAGATCTAT
ATTCCATCAACCCTACTTGTTATTGTCTCATGGGTATCATTCTGGCTAGAAAGGACAGCA
GTGCCAGCGCGAGTTACTCTCGGAGTCACAACTCTTCTTACTATGACTACACAAGCCGCC
TCCATCAACAGCAGCTTGCCAGCCGTTTCCTACGTAAAAGCAGTAGATGTCTGGATCGGA
GTTTGTCTCGCATTCATATTTGCAGCTGTGCTTGAATTTGCATGGGTTTCTTATAGAGGGT
CTTTGCTAAAAGCATGTCCAGAACATTCCCGCATGATGCACAGCAGTCCCGACACGGAC
AGCCCTGAAGCTCAGAGTTCGGAGGAAAGTAGAGTCAAACGGCCGGTAAAGAAAGACA
GTTGGGGAGGTGCAAAACACGATGAACAAGCGCTATTTGATGAACTTGAAAAACCTCTT
CGTAAATCGCCCCGAACATGGTGGGAAAAATGGAAACTCAACGCCGACCCTCCAAAAAT
GATTGATCTCAGATCGCGCGTGATCTTCCCTATACTGTTCTTGTTATTCAACATATTCTAC
TGGACGTGGTTTTCACTCAGATGAGGAATACTGACTTCATTTGAAGGGTTGTACACATTG
CTCACATGACAAATCCTCCAAGCTGGTTCCATAGAGCAAGCAATTTCGTGTCCGAGTCTT
TCTGAAAAGAAAACATAAGCAGCGAGCAAAGCGGCA

>ben-1_1
AGAAGTTCTCGTCTGCAAACATGCGTGAGATCGTGCATGTACAAGCTGGTCAGTGCGGA
AACCAGATCGGCTCAAAGTTCTGGGAAGTGATCTCCGATGAGCACGGTATCAAGCCAGA
CGGCACATACCAGGGAGAGTCCGATCTTCAATTAGAGCGAATCAATGTATACTATAACG

AAGCACATGGAGGTAAATATGTACCACGTGCAGTCCTTGTTGATCTCGAACCCGGAACT
ATGGACTCTGTCCGTTCTGGGCCATACGGACAATTGTTCCGTCCCGATAACTACGTGTTT
GGACAGTCCGGTGCAGGAAACAACTGGGCGAAGGGTCACTACACTGAAGGAGCTGAGC
TCGTTGATAATGTTCTTGATGTAGTTCGCAAGGAAGCTGAAGGATGCGACTGTCTCCAAG
GCTTCCAATTGACTCACTCGCTTGGAGGAGGTACCGGCTCGGGTATGGGAACTCTCCTCA
TCTCAAAGATCCGTGAGGAATACCCTGACAGAATCATGTCCTCATTCTCCGTGGTACCAT
CACCCAAGGTCTCGGACACTGTGGTAGAGCCATACAATGCCACTCTTTCTGTTCACCAGC
TGGTCGAGAATACAGATGAGACTTTCTGCATCGATAATGAAGCTTTGTATGATATTTGCT
TCCGCACTCTGAAACTCACAAACCCAACTTATGGAGATCTGAATCATCTCGTGTCTGTAA
CTATGTCTGGTGTAACTACTTGCCTTCGCTTCCCTGGTCAATTGAATGCTGATCTTCGTAA
ACTGGCTGTCAATATGGTTCCATTCCCTCGACTTCACTTCTTCATGCCTGGATTTGCTCCT
TTGTCTGCTAAAGGTGCTCAGGCTTACCGCGCACTTACCGTTGCTGAACTCACACAACAG
ATGTTCGATGCTAAGAACATGATGGCTGCCTGTGACCCGCGACATGGTCGATACCTCACC
GTGGCTGCAATGTTCCGTGGACGAATGAGCATGAGGGAAGTGGACGACCAAATGATGTC
TGTGCAGAACAAGAACTCCTCATACTTTGTGGAATGGATTCCGAACAACGTAAAGACCG
CTGTGTGCGATATTCCTCCACGAGGCCTGAAAATGGCAGCCACCTTTGTTGGAAACTCAA
CTGCGATCCAAGAGCTGTTCAAGCGCATCTCAGAACAATTTACTGCTATGTTCCGCCGTA
AGGCCTTCTTGCATTGGTACACCGGTGAGGGTATGGATGAGATGGAATTCACCGAGGCC
GAGTCCAACATGAATGATCTCATCTCCGAGTATCAGCAGTACCAGGAGGCTACTGCTGA
TGACATGGGCGATCTCGAGGCAGAAGGCGCTGAGGAGGCTTATCCAGAAGAGTAGACC
AACATACCAGCAACTTGTTGTGTTGTTCATTCCCTGTGTCAATGCGAAATACACATTGGT
TGCGTT

>ben-1_2
GCTCAGAGAAGTTCACCTTTTTCGTTCCACACATTCAAACTTTCCTCATCATCCATCAGAC
ACGACAACACCACTGCAGAATTTGCGTTCCATTTATGATCATCGGGAAATTCGCAAACA
AAAGATAAATTATGACAAAATTAGGAAGGTGTCAACAATGCAGATCATATATTATAAAA
GTGGCAACGACAAAAATTAGCATAAGAAACCGTGTGCGGAAAAGTTACGCAGTGATGAT
AGCTGCAAGAATGAGATGCGATTCAATACTAGTTCTGTTTCTTTACTCCTCTGCGTAGGT
CTCGTTCTCGACGGTTCCCTCCATCTCTCCCTCATCATCAGCGGTAGCTTCTTGATATTGT
TGGTACTCTGACACTAGATCGTTCATGTTCGATTCGGCCTCGGTGAATTCCATCTCATCCA
TACCCTCACCGGTGTACCAATGCAAGAAGGCTTTTCGGCGGAACATAGCAGTGAATTGTT
CCGAGATGCGTTTGAACAGTTCCTGGATGGCCGTCGAGTTTCCTACGAACGTAGCAGAC
ATCTTCAGGCCTCGTGGTGGGATATCACAGACAGCTGTCTTGACGTTATTCGGGATCCAT
TCAACAAAGTAAGAGGAGTTCTTGTTCTGCACCGACATCATCTGGTCGTCAACTTCCCTC
ATGCTCATTCGTCCACGGAACATAGCAGCGACAGTAAGGTATCGTCCGTGGCGAGGATC
GCAAGCGGCCATCATGTTCTTTGCGTCGAACATCTGCTGGGTAAGTTCTGCAACAGTAAG
CGCACGGTAGGCTTGCGCTCCCTTTGCTGACAGTGGTGCGAATCCGGGCATGAAGAAGT
GAAGTCGTGGGAATGGTACCATGTTGACGGCGAGTTTTCGCAGATCAGCATTTAACTGA
CCAGTGAAGCGGAGGCAGGTGGTAACTCCAGACATGGTTACAGACACTAAGTGGTTGAG
ATCTCCATAAGTAGGGTTGGTGAGTTTCAGTGTTCGGAAGCAGATATCATAAAGTGCTTC
GTTGTCAATGCAGAACGTCTCGTCAGTGTTCTCGACAAGTTGGTGGACGGAAAGGGTAG
CGTTGTATGGCTCGACGACGGTGTCCGAAACCTTAGGTGACGGAACAACCGAGAACGAT

GACATAATTCGATCAGGATACTCTTCACGGATTTTGGCGATCAGAAGGGTGCCCATACCT
GATCCAGTACCACCACCAAGCGAGTGCGTGAGTTGGAAGCCCTGAAGGCAATCGCAGCC
CTCAGCTTCTTTGCGGACAACATCCAAAACACTGTCTACAAGCTCTGCTCCTTCGGTGTA
GTGACCCTTGGCCCAGTTATTTCCGGCTCCAGACTGTCCAAAGACAAAGTTGTCAGGTCG
GAAGAGGGCTCCAAATGGTCCAGAGCGGACTGAGTCCATCGTTCCCGGCTCAAGATCGA
CCAAAACGGCACGTGGGACATACTTTCCTCCGCTTGCTTCATTATAGTAGACGTTTATTC
GTTCTAGCTGAAGATCTGAGTCGCCTTTGTACATTCCATCGGGTTGGATGCCATGCTCAT
CGGAGATAACCTCCCAGAACTTAGCTCCGATCTGGTTACCACACTGTCCGGCTTGAATAT
GTACGATCTCACGCATTTTGTAGAGAAGTTCGTGTCACTGACGACGAGACTGAAATAAC
GTCGTTTCGACGAAGGCTCGAAGCGGCAGCGCTGCGAAATGGGCGGGAACGCCTCGGCG
AAAAAACCAATCACAAGGCGAG

>glc-2
TGAAAAACATGTGTTGGCTCATAGTCGCCGCGTTAGCGACCGCATGTGCTGCTAGTGGTA
CCCAAGAGCAGGAGATTCTCAATGAGTTACTGCGCAATTACGATATGCGGGTTCGCCCG
CCACCTAGCAATTACTCCGATCCACACGGGCCGGTTATAGTACGTGTAAATATAATGATA
AGAATGTTATCCAAGATAGACGTAGTTAATATGGAATACAGTATGCAATTGACGTTTCGT
GAGCAATGGGTCGATTCACGACTTGCCTACTCTCAACTCGGCTACCGTAACGCACCAAG
ATTCCTTACTGTACCACATATCAAAAGTAACCTATGGATACCAGATACATTTTTCCCGAC
CGAAAAAGCTGCCCATCGACATCTTATCGACACTGACAACATGTTCCTGAGGATACATCC
GGATGGAAGGGTGTTATACAGTAGTCGGATAAGTATCACTAGCTCATGTCACATGCAGC
TTCAACTCTACCCCTTTGACTTACAGTATTGCGATTTTGACCTAGTCAGCTACGCTCACAC
AATGAAGGATATTGTCTATGAGTGGGATCCTGATACGCCAGTACAGCTAAAGCCCGGAG
TGGGTAGTGATTTACCCAACTTTCAGCTGCAGAACATTACTACCAACGATGACTGCACTA
GCCATACGAACACAGGGTCATATGCTTGCTTGAGGATGAAGCTAATACTGAAAAGGCAA
TTCAGTTACTACCTAGTCCAGCTCTACGGGCCTACCACAATGATAGTGATCGTCTCATGG
GTATCATTTTGGATCGATATGCATTCTACGGCTGGCCGAGTTGCTCTTGGCGTCACCACT
CTTCTAACCATGACAACCATGCAAGCCGCAATCAACGCGAAATTGCCTCCGGTGAGCTA
TGTTAAAGTGGTGGACGTGTGGCTAGGAGCTTGCCAAACCTTTGTCTTCGGCGCTCTACT
GGAGTACGCGTTCGTCTCCTACAAGGATACGTCACGCAAAGAAGAGCAGGCTAAGAATA
ATGCGACACGGAAAGCCACGACTCAAAAACGTCGAGCGAAGTTGGATATTCCTATGGAT
GCCTACCAGCCACCTTGTACCTGCCATTTGTACCGCGACATGCGGCCATCGTTGCCTGAC
CGCATCCGGCGCTATTTCACCAAGCCCGACTATCTGCCGGCGAAAATCGACTTCTATGCG
AGATTTTACGTGCCGCTCGGCTTCGTCGCCTTCAACATCGTCTACTGGACGTCTTGCATTA
TGATGGCCTCAAACAAGTATCACTCCTAACTCCCCCGCACTATCCTTCACGATTGCTCAC
ATTATGACATGGTTATTAAAGGCAGTTCATATCAATGATTCTGTTTTTCTGACGTGTATTG
TCTGTAATTGCGGA

>glc-3
GAAAGCAACAGCATCAGCATTGTAATGGTCGTTAGCGAGCCGTACACTTATGCACCACT
GCATCTACATTGTGTTACTTCTCATGTATTCTAATGCACTTGAAAAGAAAAAGAGTGAAT
TTTGACCAGTGCCACTATTTCGATGCGGCCACGGATAGCCGGATGGGTAGCCGTCATTCT
AGGAGTGCAGCTTTTTTGCCGCGTTTCCACCAGGCAATATCCGCGTCGACGAGAGCATGT

CCATAAGACGTCGGCCAAGGAAGACGAACAGAATGAGAACGATAGCGCCTCTTTGGAC
GAGCTCATCTACTCGATCGAAGAGAATTCCAGTCCTAGGTTGAAGCCGGCTGCTGCACCC
AGTGACGAGCAGCCGGAAGCTGTGACGGATATACCTGGGGCTCGACGAGGTGTTATTGA
AGCCGATGCCACATCAGACACCGAGATCATCAAAAAATTGCTCAACAAAGGGTACGACT
GGAGAGTACGTCCACCCGGGATCAATTTAACAGCGAAGGGAAGCCATGGACCAGTCGTA
GTTAACGTCAACATGCTCATTAGGAGCATTTCCAAAATCGACGACGTGAACATGGAGTA
CAGTGTGCAGCTGACGTTTCGAGAGTCATGGGTGGATGGCCGACTGGCCTACGGGCTCC
CTGGCGACAATAAGCCCGATTTTTTGATCCTGACAGCTGGCCAACAGATCTGGATGCCGG
ATAGCTTTTTCCAGAACGAAAAACAAGCGCAGAAGCACATGATCGACAAGCCGAATGTG
CTGATTCGAGTGCACAAAGATGGCCAAATTCTGTACTCTGTCAGGATATCTCTAGTTCTT
TCATGCCCTATGCATTTACAGTACTATCCTATGGATGTGCAAACTTGCCTGATCGACTTA
GCCTCATATGCATATACAGACAACGATATAGAATATCGCTGGAAAGAAAAAGACCCGGT
CCAGCTGAAAGATGGACTCAACTCCTCATTACCGAGCTTTCAATTGAACAAGGTGACTAC
CACATATTGCACTAGCAAAACAAATACGGGTACATATTCATGCCTGAGGACTATTTTAGA
GCTTAGGAGGCAATTTAGTTATTACCTTTTGCAATTGTACATTCCATCCTGTATGTTGGTC
ATAGTCTCATGGGTGTCCTTTTGGTTGGATCGAACGGCTGTACCCGCGAGGGTTACTCTG
GGAGTTACCACTCTGCTGACGATGACTACACAGGCATCTGGTATAAACGCCAAGTTGCC
GCCAGTATCGTATACGAAGGCTATAGACGTGTGGATTGGAGCATGTTTAACGTTCATTTT
TGGCGCTTTGCTGGAATTCGCCTGGGTTACCTATATATCGACCAGGACACAGTCTAAAAA
TACTCGGCCAGAACCGCGGACGAGTTCGCTCGTGATGTCCAATCAACATGTCATTTTACC
GCGGACAACGCTGGATTTACGAGCCCGGGGTACCGTCGACGGCGACATCTGGGTGAAGC
GAGGTCGTTTCGATGAGGCCGCGGAGTTATTGGTGCTCAGTCCGCGGCCTATTAGTCGAA
CAACTCGCTTCAAAAAGATGCTGAAGAAGTCTTGCTTGATCTCATGGGTGCGGCAACGG
CTCGAACCCGCCGATAGCGCCAAAAGGGCGGATCTTGTTTCTAGAGCCCTATTTCCGATG
TGTTTCATCCTCTTCAACATATTATACTGGACCAGCTACTCACAGTACCACGTGCCTGGA
CCGCGGTAGTACCGCAGCCTCATCAGACCTCCCCGGTTGCCCAACGGAGCAACCAATTTC
GGAAATTCCGGACGACAACGATCTTTGGAAGGAAAATGCTGATCATCAGTGGAAAACGC
ACCTCTTCACGATCTCAAAGGAAAATTCCCACCTGGTTGATCTCAGCGTGTTCTTGCCAG
CTCACCTCATTTCCACGTTCTTTCCACATTTGTGAATTTCATGATCAACATTATCATCGAA
TGTATAATTTATTGTGCATTTTCATTGTATACAGTGAAATCTCAGGCAGTTTTATATGAAA
ATTTAAATGAGAGTTATTA

>glc-4
CCGAGTAATTGTGATGCTCGCCGTCCGATGAACATCCATGCTATTCTCACAATCGCACTC
GTCTTGCTGCCATTCTCGAACTCACGAAGAGCTAGGAAGAAAGCATGTAAAAGAACTAC
CTTCGATCGACACACGACCAACTATCAGGCATGGCGGGAGCAGATGACAGTATGCGACC
TGCTGCAGGAGTACGACGCAGCCGTCAGACCAAGCGGACGCACACCGTACAACAATAGC
AAGGGTCCTGTGATAGTTACGACGAGCCTCAATATTCGATCTATTTCGGCAGTTTCAGAG
AAAAACATGGAATTTGTGGCGCAGTTTCGGTTTCGGCAAGAATGGTTCGACGACCGGCT
GCGATTCATGGATCACCAGGGCCCAATTTCTCCAGAATATCGAAACTTTGAATTTATTCA
TGTAGCTCGTGATCAGCGGTTGTGGATTCCTGACACATTTTTCCAAAACGAAAGAAACGG
GTGGTACCACATGTTGGACCAGGAGAATCGCTTTTTAAAGATTCGCTCAGACGGGAAAC
TTATTTATGATAGAAGGTTAACACTTCACTTATCTTGTTCTATGCATTTGTCACGTTACCC

AATGGATTCGCAGCTTTGCGAAATTGCTTTTGCGTCCTATGCTTACACCACTGACGATAT
AAAGTACGAGTGGGATGCTGAGGCGATCCGAATTCACGACGGAGCGAATGGAGCACTG
CCGAACTTTGACATCGCCATGTTTACCAATGGAACATGTCATTCGAAGACTAACACAGGC
GAATATTCGTGCTTACGAGTGGAGCTGAAGCTGAATCGCGTGTTCTCGTTCTTCCTACTA
CAACTGTACATCCCATCGTCGATGTTAGTTGGGGTTGCATGGGTGTCTTACTGGATTGAC
TGGAAGAGTACAGCTGCTCGCGTTCCTCTGGCTATTGTTACTTTGCTCACAATGATAACC
ACCTCGCACGCCATCAACTCGAATCTGCCTCCCGTGTCCTACGCAAAGTCGATCGACATT
TGGGTTGGAGCTTGTGTGGTCTTCATCTTCTTTTCGCTGATCGAGTATGCTGTGGTAAACT
ACATGGGAATCATGGACGAGCACAGGCAAATGAGAAAAGCAGCGTGCAATCGTAGCCG
ACTTTCAAATGTGATCGACAACCAGTTCTACATGCAATGTGTGGATCAGATGACTTCACC
CAACTCCATTATTGGGTTTAGCCCTCAAGAAAAGAAGCGACTTTTGCGTCGGAAGCGAA
ACAAATCCTTCGAACTACGCGAGGAAGAGGCTGGCATGGAGTACGATGGAATCGAACTT
GGCGAAATCGATCCTCCAAGAACAGCTGGTCACGTCGAACAAGGATGGACATTCCAGGA
CGCCACCGATCTCGTCTATATCGGTCAACGTAAAAGGGTCGAGCTCGTCCGGTGGTGTAG
TGTGCTGTCGTCTAGAGGTCGTGCAGAAAGAATAGATATAATAGCAAGGATAATCTTTC
CAATCGCCTTCATTATGTTCAACTTTGCCTACTGGAGTATATATCTGGATAATGCGGAGT
CATAGATATTCATTCCTGGTTTGTACGACTCCATCGTGCTTTCTTGTTGTCGTTGTTTTTCT
TTCAAATAAAATTTTAT

ACKNOWLEDGEMENTS

I was supported during my Ph.D. studies by a large group of people including funding agencies and non-funding assistance. I and my work were supported financially by the Hatch Act, the state of Iowa, the NIH (R01 AI047194), the George Washington Carver Fellowship, and the Multidisciplinary Graduate Education Training Fellowship. I received support from my home department of Veterinary Pathology and additional support and funding by the Interdepartmental Genetics program and the Entomology department.

My collaborators on the research chapters of this work included Dr. Richard Martin, Dr. James Koltes, Christian Bartholomay, and Jack Gallup. John VanDyk provided necessary computer assistance. I would like to thank Claude Charvet for the use of some data that was non-public at the time of use and for reviewing the manuscript that is the basis for chapter 4. Mike Baker and the ISU DNA Facility made obtaining the Illumina sequence reads used in this work extremely easy. I would like to thank my Program of Study committee for their guidance in completing my degree.

Lastly I would like to thank my family, in particular my wife, Megan, and daughter, Charlotte, who were incredibly patient and supportive through the long process that was my graduate studies.